

**Evaluation of changes in image appearance with changes in
displayed image size**

Jae Young Park

Faculty of Media, Arts and Design

This is an electronic version of a PhD thesis awarded by the University of Westminster. © The Author, 2014.

This is an exact reproduction of the paper copy held by the University of Westminster library.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch:
(<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail
repository@westminster.ac.uk

Evaluation of changes in image appearance with changes in displayed image size

JAE YOUNG PARK BSc(Hons), MSc

A thesis submitted in partial fulfilment of the
requirements of the University of Westminster
for the degree of Doctor of Philosophy

May 2014

This research programme was carried out within the Imaging Technology
Research Group at the University of Westminster

Abstract

This research focused on the quantification of changes in image appearance when images are displayed at different image sizes on LCD devices. The final results provided in calibrated Just Noticeable Differences (JNDs) on relevant perceptual scales, allowing the prediction of sharpness and contrast appearance with changes in the displayed image size.

A series of psychophysical experiments were conducted to enable appearance predictions. Firstly, a rank order experiment was carried out to identify the image attributes that were most affected by changes in displayed image size. Two digital cameras, exhibiting very different reproduction qualities, were employed to capture the same scenes, for the investigation of the effect of the original image quality on image appearance changes. A wide range of scenes with different scene properties was used as a test-set for the investigation of image appearance changes with scene type. The outcomes indicated that sharpness and contrast were the most important attributes for the majority of scene types and original image qualities.

Appearance matching experiments were further conducted to quantify changes in perceived sharpness and contrast with respect to changes in the displayed image size. For the creation of sharpness matching stimuli, a set of frequency domain filters were designed to provide equal intervals in image quality, by taking into account the system's Spatial Frequency Response (SFR) and the observation distance. For the creation of contrast matching stimuli, a series of spatial domain S-shaped filters were designed to provide equal intervals in image contrast, by gamma adjustments. Five displayed image

sizes were investigated. Observers were always asked to match the appearance of the smaller version of each stimulus to its larger reference.

Lastly, rating experiments were conducted to validate the derived JNDs in perceptual quality for both sharpness and contrast stimuli. Data obtained by these experiments finally converted into JND scales for each individual image attribute. Linear functions were fitted to the final data, which allowed the prediction of image appearance of images viewed at larger sizes than these investigated in this research.

List of Contents

Abstract	i
List of Contents	iii
List of Figures	xi
List of Tables	xviii
Acknowledgements	xx
Author's declaration	xxi
1. Introduction	1
1.1 Aim and objectives	3
1.2 Structure of thesis	4
1.3 Related publications	5
2. Image quality and appearance	6
2.1 Overview of image quality	8
2.2 Objective evaluation	10
2.2.1 Perceptual image quality attributes	11
2.2.2 Objective measures related to perceptual image quality attributes	14
2.2.2.1 Tone reproduction and contrast	14
2.2.2.2 Colour reproduction	16

2.2.2.3	Resolution	21
2.2.2.4	Sharpness	22
2.2.2.5	Noise and digital artefacts	23
2.2.3	Image quality metrics (IQMs)	26
2.3	Subjective evaluation	28
2.3.1	Overview of psychophysics and psychometric scaling	28
2.3.2	Scale types	29
2.3.3	Scaling methods	31
2.3.3.1	Threshold evaluation	32
2.3.3.2	Supra-threshold evaluation	34
2.3.4	Visual matching technique	36
2.4	Measuring and modifying sharpness	36
2.4.1	SFR evaluation	38
2.4.2	Image sharpness manipulation	40
2.4.2.1	Filtering in spatial domain	40
2.4.2.2	Filtering in frequency domain	41
2.4.2.3	Softcopy ruler method for generating sharpened and blurred images with known MTF	43

2.5	Measuring and modifying tone reproduction and contrast	47
2.5.1	Opto-Electronic Conversion Function (OECF)	47
2.5.2	Electro-Optical Transfer function (EOTF)	49
2.5.3	Formulae for contrast evaluation	50
2.5.3.1	Michelson contrast	50
2.5.3.2	Weber fraction definition of contrast	50
2.5.3.3	Root mean square (RMS) contrast	51
2.5.4	Image contrast enhancement	52
2.5.4.1	Histogram equalisation	52
2.5.4.2	Contrast stretching by piecewise linear transformation function	52
2.5.4.3	Contrast enhancement using an ‘S-shape’ function	53
2.6	Scene dependency and classification	53
2.6.1	Scene dependency	53
2.6.2	Classification of scenes	54
2.7	Appearance versus image size	55
3.	Device characterisation	57
3.1	Digital cameras	59

3.1.1	Tone characteristics (Opto-Electronic Conversion Function).....	60
3.1.2	Colorimetric characteristics of sRGB output.....	63
3.1.3	SFR measurements using the slanted edge method	67
3.1.4	Summary	70
3.2	Liquid crystal displays (LCDs)	72
3.2.1	Conditions of measurement, calibration and settings	72
3.2.2	Tone characteristics (Electro-Optical Transfer Function).....	74
3.2.3	Basic colorimetric characteristics	76
3.2.4	Colour tracking characteristics	78
3.2.5	Positional non-uniformity	82
3.2.6	Dependency on background	87
3.2.7	Temporal stability	87
3.2.8	Viewing angle dependency	90
3.2.9	Positional non-uniformity at the observation plane	94
3.2.10	Summary	95
4.	Psychophysical investigation 1: Identification of image attributes	
	that are most affected with changes in displayed image size	98
4.1	Preparation of test stimuli	99

4.1.1	Image capture	99
4.1.2	Image selection	100
4.1.3	Image processing	100
4.2	Psychophysical investigation	101
4.2.1	System calibration and settings	102
4.2.2	Software preparation and interface design	102
4.2.3	Rank order method	103
4.3	Classification of test images	104
4.4	Results and discussion	106
4.5	Summary	114
5.	Psychophysical investigation 2: Evaluation of changes in perceived sharpness with changes in displayed image size	117
5.1	Preparation of test stimuli	118
5.1.1	System tone reproduction	118
5.1.2	System SFR	119
5.1.3	Determination of the reciprocal measure of the system bandwidth, k	121
5.1.4	Sharpness filters	123

5.1.5	Frequency domain filtering and bi-cubic interpolation	125
5.1.6	Effect of bi-cubic interpolation on image quality	125
5.2	Psychophysical investigation	127
5.2.1	Display settings and calibration	127
5.2.2	Software preparation and interface design	128
5.2.3	Sharpness matching experiment	128
5.3	Results and discussion	130
5.3.1	Results from the psychophysical tests	130
5.3.2	Validation of the results	134
5.3.3	Evaluation of step interval and calibration of changes in sharpness JND scales	135
5.4	Summary	137
6.	Psychophysical investigation 3: Evaluation of changes in perceived contrast with changes in the displayed image size	139
6.1	Introduction	141
6.2	Preparation of test stimuli	142
6.2.1	Creation of a series of contrast filters with n-JND interval	142
6.2.2	Spatial domain filtering	145

6.2.3 Contrast measurement of the ruler images	146
6.3 Psychophysical investigation	148
6.4 Results and discussion	149
6.4.1 Results from the psychophysical tests	149
6.4.2 Validation of the results	153
6.4.3 Evaluation of step interval and calibration of changes in contrast JND scales	153
6.5 Summary	155
7. Discussion	157
7.1 Capturing devices	157
7.2 Display devices	159
7.3 Identification of image attributes	160
7.4 Sharpness matching	161
7.5 Contrast matching	163
8. Conclusions and recommendations for further work	165
8.1 Conclusions	165
8.2 Recommendations for further work	167
Appendices	168

A. Thumbnails of test images	168
A.1 16 average scenes	168
A.2 Test images (in alphabetical order)	169
B. Instructions for observers	172
B.1 Observer instructions for rank order experiments	173
B.2 Observer instructions for sharpness matching experiments	174
B.3 Observer instructions for contrast matching experiments	175
B.2 Observer instructions for result validation experiments	176
B.2 Observer instructions for step validation experiments	177
C. Publications	178
D. List of abbreviations	210
References	213

List of Figures

2-1.	The relative importance of the FUN dimensions on the quality of different image types	13
2-2.	The triangle of colour	17
2-3.	Visually equal chromaticity steps at constant luminance on CIE 1931 x, y diagram (left) and some of the steps re-plotted in CIE 1976 u', v' diagram ...	18
2-4.	Three-dimensional representation of the CIELAB L^* , a^* , and b^* coordinates	20
2-5.	Measures or models describing the images' or the imaging systems' attributes and models of the HVS are used in IQMs	26
2-6.	Illustration of psychometric scales	31
2-7.	A typical psychometric curve	33
2-8.	Flowchart of the deviation of digital SFRs from captured slanted edges	39
2-9.	Examples of commonly used spatial domain linear filters. Blur filters (left top, left bottom) and Laplacian sharpening filters (right top, right bottom)	41
2-10.	The imaging equation (convolution) and the spatial frequency equivalent	41
2-11.	Number of operations required to perform convolution in spatial and frequency domains on a 1024×1024 pixel image versus kernel size	42

2-12.	Perspective plots of a Butterworth lowpass filter (left), and a Gaussian highpass filter (right) transfer functions with their images	43
2-13.	Plot of Equation 2.16, spaced by 3 JNDs (left) and Equation 2.17 (right)	44
2-14.	Implementation of steps involved in the creation of frequency domain Gaussian filters, with a constant interval based on ISO 20462-3	46
2-15.	Two different test charts for measuring transfer functions of acquisition devices	48
2-16.	Typical electro-optical transfer functions for CRT and LCD devices	49
3-1.	Tone reproduction characteristics of the Apple iPhone camera (top) and the Canon 30D camera (bottom)	62
3-2.	Original and the captured red, green, blue, and white patches of the GretagMacbeth™ ColorChecker by both cameras	64
3-3.	Colour reproduction errors between the original and captured patches for both camera systems using two commonly used colour difference formulae	66
3-4.	Horizontal (top) and vertical (bottom) SFR of the Apple iPhone camera	68
3-5.	Horizontal (top) and vertical (bottom) SFR of the Canon 30D camera	69
3-6.	Tone characteristics of the EIZO CG210 display.....	75
3-7.	Tone characteristics of the EIZO CG245W display.....	75
3-8.	Reproduction of the full on primaries and the white on display devices and their corresponding values in sRGB colour space	77

3-9.	Colour tracking characteristics of the EIZO CG210 before (top) and after the black level compensation (bottom)	80
3-10.	Colour tracking characteristics of the EIZO CG245W before (top) and after the black level compensation (bottom)	81
3-11.	Positions of 25 selected points for positional non-uniformity characteristic of a display device	82
3-12.	Lightness differences, ΔL^* , from the reference point to the measured points across the screen	84
3-13.	Chromatic differences, ΔC_{ab}^* , from the reference point to the measured points across the screen	85
3-14.	Colour differences, ΔE_{ab}^* , from the reference point to the measured points across the screen	86
3-15.	Short-term stability in luminance (top) and in chromaticities (bottom), on the CG210 (left) and on the CG245W (right)	89
3-16.	Mid-term stability in luminance (top) and in chromaticities (bottom), on the CG210 (left) and on the CG245W (right)	89
3-17.	Luminance output of the pure primaries and the white at various horizontal and vertical viewing angles	91
3-18.	Changes in chromaticities at various viewing angles	92

3-19.	Changes in luminance output of neutral patches at various horizontal and vertical viewing angles	93
3-20.	Colour differences, ΔE_{ab}^* , from the reference point to the measured positions across the screen	95
4-1.	Display interface for the psychophysical test page in achromatic mode	103
4-2.	Average ranks from all test stimuli	108
4-3.	Average ranks of the image attributes of test stimuli categorised by their average lightness	109
4-4.	Average ranks of the image attributes of test stimuli categorised by their colourfulness	110
4-5.	Average ranks of the image attributes of test stimuli categorised by their busyness	111
4-6.	Average ranks of the image attributes of test stimuli categorised by their sharpness	112
4-7.	Average ranks of the image attributes of test stimuli categorised by their noise level	113
5-1.	Transfer function of the Camera-Display combined system	119
5-2.	Spatial frequency responses (SFRs) of the combined system at major aperture stops	120
5-3.	Modelled MTF curves with the various k values	122

5-4.	Secondary standard quality value at $k=0.030$ and $k=0.047$	122
5-5.	Cross section of blurring filters for the images taken at $f/11$ and below	124
5-6.	Cross section of sharpening filters for the images taken at $f/11$ and below	124
5-7.	Effect of the bi-cubic interpolation on SFR, Tate Modern scene	126
5-8.	Effect of the bi-cubic interpolation on SFR, Pembroke lodge sign scene	126
5-9.	Display interface of sharpness matching test with a slider	128
5-10.	Average perceived loss in image quality from the small vs. large experiment for each scene with SEM	131
5-11.	Average perceived loss in image quality from the medium-small vs. large experiment for each scene with SEM	131
5-12.	Average perceived loss in image quality from the medium vs. large experiment for each scene with SEM	132
5-13.	Average perceived loss in image quality from the large-medium vs. large experiment for each scene with SEM	132
5-14.	Perceived changes in image quality with respect to the changes in displayed image size (blue) and predicted changes (red) in non-calibrated relative image quality JND scale (SQS_2)	134
5-15.	Average ratings of the original pairs and the sharpness matched pairs	135

5-16	Changes in perceived sharpness with respect to the changes in displayed image size (blue) and predicted changes (red) in sharpness JND scale	136
5-17.	Average ratings for the sharpness modified and unmodified image pairs	151
6-1.	Sample S-shaped filter functions, calculated by gamma adjustment by power transformation	143
6-2.	A series of gamma increasing filter functions	144
6-3.	A series of gamma decreasing filter functions	144
6-4.	Sample S-shaped filters and the contrast manipulated images	145
6-5.	C_{RMS} of four selected scenes at a different ruler scale	147
6-6.	C_{RMS} of ‘Regent’s Park 2’ at a different ruler scale in 3 different image sizes	147
6-7.	Average perceived change in tone reproduction from the small vs. large experiment for each scene with SEM	150
6-8.	Average perceived change in tone reproduction from the medium-small vs. large experiment for each scene with SEM	150
6-9.	Average perceived change tone reproduction from the medium vs. large experiment for each scene with SEM	151
6-10.	Average perceived change in tone reproduction from the large-medium vs. large experiment for each scene with SEM	151

6-11.	Perceived changes in tone reproduction with respect to the changes in displayed image size (blue) and predicted changes (red) in non-calibrated relative image quality gamma scale	152
6-12.	Average rating of the unmodified pairs and the contrast modified pairs	153
6-13.	Changes in perceived contrast with respect to the changes in displayed image size (blue) and predicted changes (red) in contrast JND scale	154

List of Tables

2-1.	Image attributes examined in image quality assessment and associated perceptual attributes	11
2-2.	Categorisation of selected image quality attributes	12
2-3.	Imaging performance measures relating to the objective evaluation of imaging systems	14
2-4.	Colour attributes and definitions	21
2-5.	Noise in image sensors	24
2-6	Common digital image artefacts, their sources, and areas within images which are more susceptible to those artefacts	25
2-7.	Stevens' classification of scale types	30
3-1.	Camera settings for the image capture	59
3-2.	Colour differences between the original and captured patches for both camera systems	65
3-3.	Technical specifications of display devices and the settings used during calibration and experiments	73

3-4.	CIE 1931 tristimulus values and CIE 1976 chromaticity coordinates for the pull on primaries and the white from both display devices	76
3-5.	Measured CIELAB values and evaluated colour differences	87
4-1.	Images classified according to their lightness, colourfulness, busyness, sharpness and noisiness	106

Acknowledgements

I would like to express my sincere thanks to my supervisor Dr. Sophie Triantaphillidou, director of my studies, and my second supervisor Professor Ralph Jacobson of University of Westminster, for the original idea for the project, their constant guidance and encouragement throughout the duration of this research.

I received much help and inspiration from colleagues in the group. In particular, thanks are due to Dr. Gaurav Gupta, Anastasia Tsifouti, Moacir Lopes, Edward Fry, and Kyung Hoon Oh for many useful discussions, encouragement and friendship. I also thank Dr. John Jarvis, Dr. Efthimia Bilissi, Dr. Olivier Moulard and Elizabeth Allen for useful discussions and advice.

Special thanks go to all those who participated as observers in my experiments.

Finally and most importantly, this thesis is dedicated to my wife and my family for their unflagging support.

Author's declaration

I declare that all the material contained in this thesis is my own work.

Chapter 1

Introduction

Developments in the last few decades in display technology have replaced CRTs with new displays, such as liquid crystal displays (LCDs), plasma display panels (PDPs) and organic light-emitting diodes (OLEDs), with the LCDs still being the most popular technologies for viewing computer images. LCD devices are not restricted to computer monitors and domestic televisions, but also to many modern and mobile devices (i.e. digital cameras, mobile phones, and etc.). Nowadays, digital images are increasingly viewed in various sizes using different display devices. The advancement in display technology has resulted in satisfactory image reproduction on built-in LCDs in

capturing devices under various viewing conditions. Since the quality of image reproduction on such LCDs is often satisfactory, camera users can assess thumbnails of captured images displayed on the built-in LCDs immediately after capture to judge overall image quality.

However, such judgements made about captured images on LCDs are often incorrect. For example, images viewed on small displays are likely to appear much sharper in most cases than when they are viewed at a larger magnification on a computer display. The properties, physical and pixel sizes of small camera displays affect the way the images are displayed, which in turn produced distorted visual information about the captured images. Image appearance is a phenomenon of visual perception. Subjective impressions of image quality are naturally affected by various factors including the surrounding viewing conditions for images, the physical changes in image size, and the changes in the angle subtending the observer's eye (Choi *et al.*, 2007b, Nezamabadi and Berns, 2006, Nezamabadi *et al.*, 2007, Xiao *et al.*, 2010, Wang and Hardeberg, 2012, Xiao *et al.*, 2011).

Research has been carried out to identify and quantify changes in image appearance with respect to image size and viewing angle. However, most of relevant studies were conducted using uniform colour patches, or artificially generated test patterns. In colour appearance investigations by Choi *et al.* (Choi *et al.*, 2007b), the studies mainly focused on changes of the size of uniform patches, illumination levels, and surround and relative display luminance. Measurements in these studies were restricted to a certain area of the display with a small coverage under certain viewing conditions. In another study, Nezamabadi *et al.* (Nezamabadi and Berns, 2006) investigated changes in perceived lightness and chroma with changes in visual angle

and thus perceived image size. Also, Xiao *et al.* (Xiao *et al.*, 2010, Xiao *et al.*, 2011) investigated the effect of image size on colour appearance. Further, Nezamabadi *et al.* (Nezamabadi *et al.*, 2007) investigated the relationship between changes in image size and perceived contrast, using contrast matching techniques and artificially generated noise patterns of different spatial frequency content. Although significant spatial effects such as sharpness, noisiness, and most importantly, the appearance of digital image artefacts caused by varying image size, image content, and illumination conditions have not been considered in depth. More recently Wang *et al.* (Wang and Hardeberg, 2012) conducted an investigation of changes in appearance of all image attributes as well as compression with changes in visual angles. They found that although hue appearance was not affected with changes in visual angle, sharpness, noisiness, and compression artefacts were affected significantly.

1.1 Aim and objectives

The aim of this research is to predict the changes in image appearance when images are viewed at different sizes on LCD devices. A study concerning the quantification of perceived changes in the two most affected image attributes with changes in displayed image size was carried out using matching techniques, a large number of scenes and selected observers.

The objectives of the research are:

- to identify the two most important image attributes, the appearance of which is most affected by changes in displayed image size;

- to quantify the changes in image appearance with changes in displayed image size in Just Noticeable Differences (JNDs) in secondary standard quality scales (SQS_2);
- to express these JNDs in quality scale to calibrated JNDs in relevant individual attribute scales.

These objectives were achieved by a series of psychophysical investigations. Initially, an investigation identified which image quality attributes were most affected by changes in displayed image size. The identified image attributes were then investigated further to quantify the degree of change in perceptual sharpness and contrast, with respect to changes in displayed image size, by matching experiments. The last step was converting the data obtained by these experiments into JND scales for each individual image attribute.

1.2 Structure of thesis

An overview of image quality is included in Chapter 2. In addition, this chapter presents common methods used in psychophysical evaluations and detailed descriptions of methods employed for the subjective evaluation of image appearance changes. Further, digital image manipulation techniques related to aspects of this research project are described. This chapter concludes with a brief introduction of objective image quality measure with respect to scene content and characteristics.

Characterisation of devices used for image capture and image display were carried out to understand the effect and limitation of the device characteristics employed in this project. Details on tone reproduction and colorimetric characterisation of the

capturing and display devices are provided in Chapter 3. In addition, Spatial Frequency Response (SFR) measurements of the capturing devices are also provided.

Chapter 4 presents a detailed description of experimental methods for investigating which image quality attributes are most affected by changes in displayed image size and presents relevant results and conclusions. Chapter 5 is concerned with the quantification of changes in perceived sharpness with respect to changes in displayed image size, which was achieved by a sharpness matching experiment. This chapter also provides a detailed description of the novel method for creating a range of test stimuli with varying sharpness levels, by taking into account the SFR of the imaging system. In Chapter 6, the experimental methods for quantifying changes in perceived contrast with respect to changes in displayed image sizes by contrast matching is given. A detailed description of the method for contrast manipulation is also provided in this chapter. In addition, the evaluation of step intervals and conversion of the results from the derived relative quality scales to univariate JND scales for sharpness and contrast are presented in Chapters 5 and 6.

Chapter 7 discusses the effects and limitations of device characteristics with respect to the psychophysical experiments carried out in this research project. In depth discussion on the results from the psychophysical investigations described in Chapters 4, 5, and 6 are also included. In Chapter 8, conclusions are drawn and recommendations for further work are proposed.

1.3 Related publications

The following related papers were produced by the author during the production of this work. Copies of them are attached in Appendix C.

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., “Identification of image attributes that are most affected with changes in displayed image size”, Proc. SPIE Image Quality and System Performance VI, 7242, 18-22 January 2009, San Jose, USA

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., Gupta, G., “Evaluation of perceived image sharpness with changes in the displayed image size”, Proc. SPIE Image Quality and System Performance IX, 8293, 22-26 January 2012, San Francisco, USA

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., “Just noticeable differences in perceived image contrast with changes in the displayed image size”, Proc. SPIE Image Quality and System Performance XI, 9016, 2-6 February 2014, San Francisco, USA

Chapter 2

Image quality and appearance

This chapter is concerned with definitions and theories related to the study and evaluation of image quality and its attributes. Factors affecting image quality and image appearance are discussed. Common methods used in psychophysical evaluations of image stimuli are provided and detailed descriptions of methods employed for the subjective evaluation of image appearance changes are presented. These methods have been applied in experimental work described in Chapters 4, 5 and 6. Objective methods used in the evaluation of imaging system performance are also presented, along with digital image manipulation techniques related to aspects of this project.

2.1 Overview of image quality

Image quality, image distortion, and image fidelity are three different aspects of the general expression ‘image quality’ and are all concerned with the assessment of images, or imaging systems (Ford, 1997, Triantaphillidou, 2001).

Image distortion is concerned with physical differences between a rendered image and an original scene or image. Distortion may occur in every step within the imaging chain or the image processing. It is evaluated numerically (objectively), using distortion measures and metrics. However, the related measures do not always have perceptual meaning or significance when the degree of distortion is imperceptible or acceptable. Therefore, results obtained from objective distortion assessments do not often correlate with perceived image quality. There are various distortion metrics commonly used such as mean squared error (MSE), root mean square error (RMSE), and signal-to-noise-ratio (SNR) (Wang and Bovik, 2002, Chapter 2 of Jain, 1989). Also, colour differences, such as CIELAB ΔE_{ab}^* and CIEDE2000 ΔE_{00}^* , can be used for measuring colorimetric distortions in CIELAB colour space. Although such colour difference are calibrated to produce results that are visually meaningful for uniform colour patches, CIELAB image differences do not always correlate with the colour appearance of images. More sophisticated appearance measures, such as image appearance models are developed for the evaluation of image appearance (Fairchild and Johnson, 2004, Kuang *et al.*, 2007).

Image fidelity is concerned with the perceptually accurate rendition (reproduction) of the original image (or original scene) (Farrell, 1999). Unlike image distortion, image fidelity assessment involves the HVS since is concerned with relative thresholds (i.e. the minimum change or difference in the images that can be visually

detected). Relative thresholds can be determined by psychophysical experiments and the results from such experiments can be used to define the just noticeable difference (JND) and the JND increment (Keelan, 2002, p.36). Details on psychophysical experiments are described in Section 2.3.

Image fidelity should be distinguished from image quality, since high image fidelity does not always imply high image quality. For example, a sharpened reproduction of a photograph that includes a lot of fine detail is often assessed to be of a higher quality than a high quality original, but quantitatively it is considered a ‘distorted, low fidelity’ version. Another example discussed widely in the literature is the portrait. A slightly blurred reproduction of a portrait is often assessed to be of a higher quality than the sharper original, since blurring provides a softer skin (Granger and Cupery, 1972). The other example is a slightly noisy reproduction of a blur original. A slightly noisy reproduction is often perceived to be of higher quality than the blur original (Cambridge in colour, 2013b).

Image quality, in strict definition, is concerned with the subjective impression of goodness the image conveys (Triantaphillidou, 2001). It describes the perceptual response of an observer to an image (or a single attribute) by taking into account the purpose of the image and psychological effect (Yoshida, 2006, p.278). When an image is viewed, observers are able to judge image quality almost instantly whether the particular image is good or poor quality. However, to quantify how good an image is, and scale its quality is a difficult operation, as Jacobson has pointed out (Jacobson, 1993).

Unlike image fidelity, image quality judgement involves the observers’ own criteria according to their personal preferences. Although image fidelity is an important

factor influencing the image quality judgement, observers take into account the purpose, or context for which the image is being used and therefore the same image may be judged differently by different observers, or under different context and conditions (Yendrikhovskij, 2002, Bilissi, 2004). Also, image quality is judged based on the observer's experience of viewing images and various other cognitive factors such as memory, emotions, influence, expectations and many more. These factors result in a variation of the assessments between individuals and temporally for the same individual (Triantaphillidou, 2001, p.32, Keelan, 2002, p.5).

Image quality is inherently a subjective attribute. However, it is measured using both objective and subjective methods. Objective methods involve measures that *ideally* correlate with the subjective impression of images. Subjective methods involve psychophysical experiments that employ human observations and statistical analysis of the results to quantify quality from qualitative assessments.

2.2 Objective evaluation

Objective evaluation of image quality involves the assessment of a number of different image quality attributes associated, in some ways, with the visual perception of images. Associated measurements are based on the assumption that there is a functional relationship between the subjective impression of image quality and selected image quality attributes of the observed image (Lockhead, 1992). It is important to identify the factors affecting the judgement of image quality and the related objective measures. In this section, image quality attributes and a short overview of objective measures used for their evaluation are presented.

2.2.1 Perceptual image quality attributes

There are various perceptual attributes (also referred to as image quality dimensions), which are related to image quality, or imaging system evaluation. Based on Miyake's work (Miyake *et al.*, 1984), Ford (Ford, 1997) listed five basic attributes along with their visual descriptions. These were originally considered for conventional analogue imaging systems but they are also valid for digital imaging systems as well. These attributes are tone reproduction (and contrast), colour, resolution, sharpness, and noise. They are presented in Table 2-1.

Image attribute	Visual description
Tone/Contrast	Macroscopic contrast, or reproduction of intensity
Colour	Differences in <i>lightness</i> , <i>chroma</i> and <i>hue</i>
Resolution	Discrimination of fine detail
Sharpness	Microscopic contrast, or reproduction of edges
Noise	Random and non-random spurious information

Table 2-1. Image attributes examined in image quality assessment and associated perceptual attributes, adapted from Ford (Ford, 1997, p.32).

Although these attributes are traditionally associated with the perception and evaluation of images and apply to all imaging systems, they are not independent from each other. This makes the subjective assessment of an individual attribute more difficult than the assessment of the overall image quality (Bartleson, 1982, Higgins and Wolfe, 1955).

More recently, there have been a number of classification approaches to image quality attributes. Keelan (Keelan, 2002, p.8) separated image attributes, of which their presence always degrade quality (*artefactual*), to *preferential* attributes that may

influence image quality, but the relationship between them and quality is not monotonic. Both artefactual and preferential attributes are related to the five basic attributes in Table 2-1. He further listed a number of attributes relating to image aesthetic and to observer preference. Keelan's approach is particularly useful for designing imaging system. For example, by assigning higher weightings to attributes in artefactual and preferential categories and lower weightings to those in aesthetic and personal categories, the fidelity of the imaging system would influence the image quality rating significantly, but it would not be the only factor. His classified attributes are presented in Table 2-2.

Category	Attribute
Artefactual	Unsharpness
	Graininess
	Redeye
	Digital artefacts
Preferential	Colour balance
	Contrast
	Colourfulness (saturation)
	Memory colour reproduction
Aesthetic	Lighting quality
	Composition
Personal	Preserving a cherished memory
	Conveying subject's essence

Table 2-2. Categorisation of selected image quality attributes, adapted from Keelan (Keelan, 2002, p.8).

Another approach is made by Yendrikhovskij (Yendrikhovskij, 2002). Author presented the FUN model (Fidelity, Usefulness, and Naturalness) of image quality, which uses three *cognitive dimensions* for the determination of quality. FUN is a modified version of his previous model, the GUN model (Genuineness, Usefulness, and Naturalness) (Yendrikhovskij, 1999). In the newer version, the author has introduced the ‘Fidelity’ as a replacement of ‘Genuineness’. As discussed in the previous section, and supported by Yendrikhovski, fidelity is concerned with the accurate rendering of image/scene and it is highly related to the attributes listed in Table 2-1. Yendrikhovski defined usefulness and naturalness attributes as below;

- Usefulness: the degree of apparent suitability of the reproduced image to satisfy the corresponding task
- Naturalness: the degree of apparent match between the reproduced image and internal reference

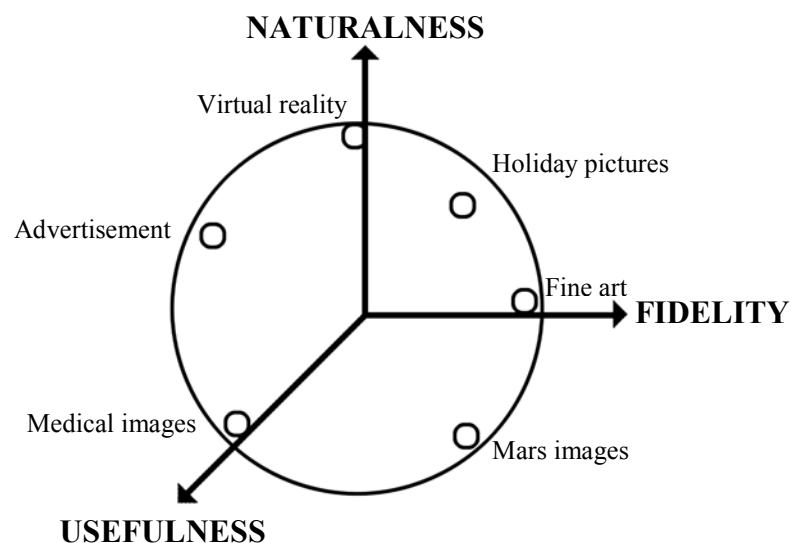


Figure 2-1. The relative importance of the FUN dimensions on the quality of different image types, adapted from Yendrikhovskij (Yendrikhovskij, 2002).

Similar to Keelan's approach, each of FUN attributes would have different weightings in different applications. Overall image quality can be modelled as a weighted sum of the three FUN attributes. Figure 2-1 illustrates the relative importance of these dimensions.

2.2.2 Objective measures related to perceptual image quality attributes

In this section, commonly employed objective image quality measures, (also referred to as imaging performance measures) which are used for the quantification of the perceptual image quality attributes, are discussed. Triantaphillidou (Triantaphillidou, 2011a) has summarised the image quality attributes and the related imaging performance measures. A summary is presented in Table 2-3.

Image attribute	Measures
Tone	Characteristic curve, density differences, OECF and EOTF, contrast, gamma, histogram, dynamic range
Colour	Spectral power distribution, CIE tristimulus values, colour appearance values, CIE colour differences
Resolution	Resolving power, imaging cell, limiting resolution
Sharpness	Acutance, ESF, PSF, LSF, MTF, SFR
Noise	Granularity, noise power spectrum, autocorrelation function, total variance (σ_{TOTAL}^2)

Table 2-3. Imaging performance measures relating to the objective evaluation of imaging systems, adapted from Triantaphillidou (Triantaphillidou, 2011a, p.349).

2.2.2.1 Tone reproduction and contrast

Tone and contrast are critical aspects of the quality of images and are related to each other. Tone reproduction describes the relationship between input and output intensities

in imaging systems; the relationship is often plotted as a *transfer function*. Contrast describes the difference in macroscopic intensities of two different areas of an image and is often expressed with a *metric* (value). Contrast is discussed more in depth in Section 2.5.

In conventional chemical imaging, the transfer function relates the output density to the logarithm of the relative input exposure and it is known as the characteristic curve. The slope (or gradient) of the linear portion of the characteristic curve is expressed as gamma, γ , and it a metric relating the mid-tone contrast reproduced by the system (Hurter and Driffield, 1890).

In digital imaging, the transfer relationship is often plotted in linear units and it is described by power function, where the exponent represents gamma having the same meaning as above. For example, the transfer function of a CRT display device can be described using the gain-offset-gamma model (also known as GOG) as Equation 2.1, adapted from Giorgianni and Madden (Giorgianni and Madden, 2008, p.33);

$$L = gPV^\gamma + o \quad (2.1)$$

where L represents the normalised output luminance on a computer controlled display, PV is normalised input pixel value, g is the system gain, and o is system offset.

The transfer function of the overall imaging chain, describing the tone reproduction of all imaging components combined together, is the product of the individual component transfer functions (Jones, 1921). A gamma correction is often applied to obtain a desirable overall system transfer function and a desired gamma. A real imaging system, including image capture and display is presented in Equation 2.2

for calculating the required gamma correction (processing gamma). The subscripts o , c , p and d represent overall, image capture, processing, and display, respectively.

$$\gamma_o = \gamma_c \times \gamma_p \times \gamma_d \quad (2.2)$$

Tone reproduction was first classified into objective and subjective tone reproduction by Jones (Jones, 1931). These terms were then formalised by Nelson in the 1960's (Nelson, 1966). Higgins later described two types of optimum tone reproduction (Higgins, 1977). One is objective tone reproduction, where optimum reproduction is achieved when gamma is equal to one, indicating that reproduced contrast is equal to scene contrast. The other is subjective tone reproduction, which takes into account the viewing conditions (Bartleson and Breneman, 1967a, Bartleson, 1975). The aim of tone reproduction is to achieve a linear reproduction of lightness (or relative brightness) and thus takes viewing conditions into account. This suggests that the optimum gamma is scene dependent (Roufs, 1989) and also influenced by the viewing conditions. Optimum gamma has been found in most imaging applications to be greater than one (c.f. Section 2.5) (Roufs, 1989, Bartleson and Breneman, 1967b, Hunt, 2004, p.92).

2.2.2.2 Colour reproduction

The perception of the colour of objects is a function of the physical properties of the objects, the light source that illuminates and the human visual system. Fairchild (Fairchild, 1998) has described three components in a form of triangle which is presented in Figure 2-2.

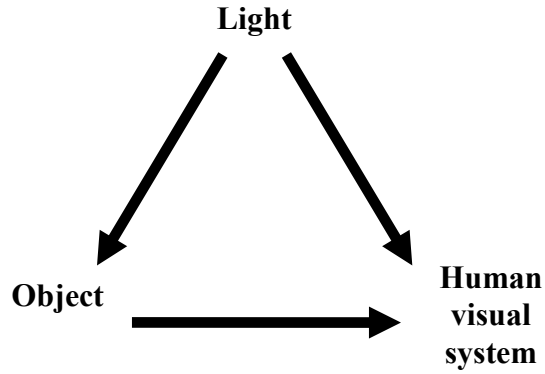


Figure 2-2. The triangle of colour, adapted from Fairchild (Fairchild, 1998, p.65).

The objective measurement and evaluation of colour reproduction is traditionally achieved by colorimetry. It is based on the theory of trichromatic vision developed by Maxwell, Young, and Helmholtz in the 19th century (Maxwell, 1871, Young, 1802). It involves the trichromatic analytical process of the spectral sensitivities of the cones in the HVS.

In colorimetry, these sensitivities were represented by the colour matching functions of the standard colorimetric observer \bar{x} , \bar{y} , \bar{z} and established by the International Commission on Illumination (CIE) in 1931 (*CIE Standard colorimetric observers*. 1991).

CIE has also established the non-physical tristimulus values, which are calculated using the \bar{x} , \bar{y} , \bar{z} colour matching functions, presented in Equations 2.3 to 2.5.

$$X = \int_{380}^{780} R(\lambda)I(\lambda)\bar{x}(\lambda)d(\lambda) \quad (2.3)$$

$$Y = \int_{380}^{780} R(\lambda)I(\lambda)\bar{y}(\lambda)d(\lambda) \quad (2.4)$$

$$Z = \int_{380}^{780} R(\lambda)I(\lambda)\bar{z}(\lambda)d(\lambda) \quad (2.5)$$

where 380 to 780 nm is the range of wavelengths, λ , of the visible spectrum, $R(\lambda)$ is the spectral illuminance, reflectance (or transmittance) of the object, or substance, and $I(\lambda)$ is the absolute or relative spectral power distribution of the selected illuminant.

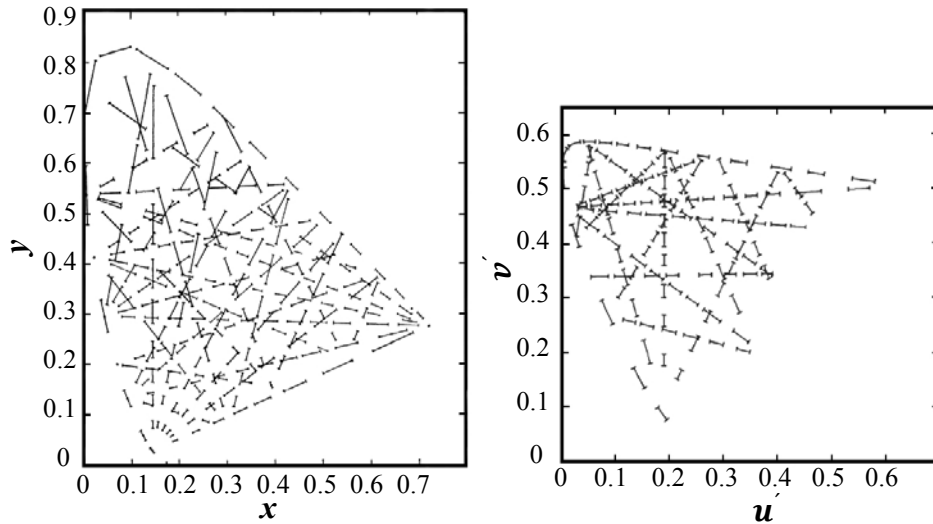


Figure 2-3. Visually equal chromaticity steps at constant luminance on CIE 1931 x, y diagram (left) and some of the steps re-plotted in CIE 1976 u', v' diagram, adapted from Hunt (Hunt, 2004, p.105-106).

The tristimulus values can be used to calculate various chromaticity coordinates, such as the x, y, z for the CIEXYZ 1931 system and the u', v', w' for the more perceptually uniform the CIELUV 1976 system. Chromaticity coordinates are plotted in 2D chromaticity diagrams as shown in Figure 2-3.

However, these chromaticity diagrams do not provide any luminance information and thus provide incomplete information on colours. For a complete colour specification, CIE has defined and recommended two uniform colour spaces, the CIELAB and the CIELUV, which employ common lightness, L^* . In this research, the CIELAB colour space was used for various image processing steps. The full coordinates

for the CIELAB colour space can be calculated by a non-linear transformation of the CIE 1931 XYZ tristimulus values, using Equations 2.6 to 2.9

$$L^* = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \quad \text{for } \frac{Y}{Y_n} > 0.008856 \quad (2.6)$$

$$L^* = 903.3 \left(\frac{Y}{Y_n} \right) \quad \text{for } \frac{Y}{Y_n} \leq 0.008856$$

$$a^* = 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \quad (2.7)$$

$$b^* = 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right] \quad (2.7)$$

$$C_{ab}^* = \sqrt{(a^*)^2 + (b^*)^2} \quad (2.8)$$

$$h_{ab} = \tan^{-1} \left(\frac{b^*}{a^*} \right) \quad (2.9)$$

where X_n , Y_n and Z_n are the tristimulus values for the reference white. The colours are defined by L^* (lightness), a^* (red-green component), b^* (yellow-blue component), C_{ab}^* (chroma), and h_{ab} (hue angle).

A three-dimensional representation of the CIELAB L^* , a^* , and b^* coordinates is shown in Figure 2-4 (Fairchild, 2005).

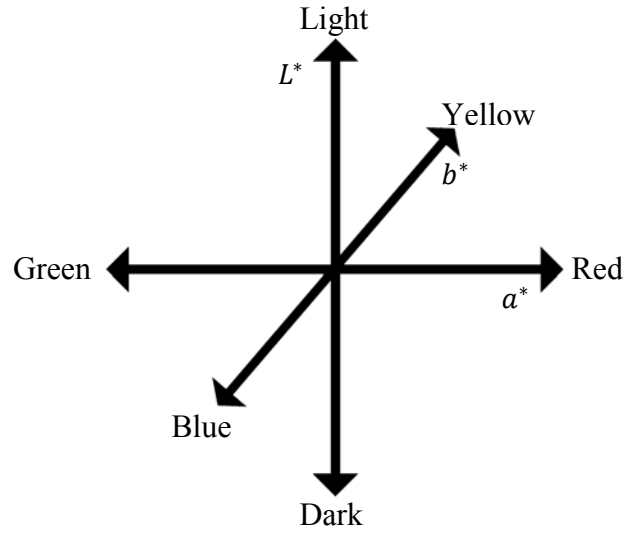


Figure 2-4. Three-dimensional representation of the CIELAB L^* , a^* , and b^* coordinates, adapted from Fairchild (Fairchild, 2005, p.80).

Colour reproduction of the imaging system is usually evaluated by some well-known colour difference models, such as CIELAB ΔE_{ab}^* , and CIELUV ΔE_{uv}^* , using the coordinates described above. The CIELAB ΔE_{ab}^* equation is presented in Equation 2.10.

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (2.10)$$

However, such colour difference models are not concerned with various issues related to the appearance of colour stimuli. Various factors affect the visual appearance of colours, including visual adaptation (light, dark, chromatic), the background and surrounding colours, and the luminance level. A number of new colour difference and appearance models, such as the CIEDE2000 (designed for uniformed stimuli), *i*CAM, and *i*CAM06 (designed for image stimuli) have been developed to represent the appearance of colour numerically. They are currently being used for the evaluation of perceptually meaningful colour differences (Fairchild and Johnson, 2004, Kuang *et al.*, 2007, Luo *et al.*, 2001).

This research is based on the visual assessment. Colour attributes, which are used in colour evaluation in latter chapters, are presented in Table 2-4. These were well defined by the CIE (*ILV: International Lighting Vocabulary*. 1987).

Attribute	Definition
Hue	Attribute of a visual sensation according to which an area appears to be similar to one of the perceived colours: red, yellow, green, and blue, or to a combination of two of them.
Colourfulness	Attribute of a visual sensation according to which the perceived colour of an area appears to be more or less chromatic.
Chroma	Colourfulness of an area judged as a proportion of the brightness of a similarly illuminated area that appears white or highly transmitting.
Saturation	Colourfulness of an area judged in proportion to its brightness.
Brightness	Attribute of a visual sensation according to which an area appears to emit more or less light.
Lightness	The brightness of an area that appears to be white or highly transmitting.

Table 2-4. Colour attributes and definitions, adapted from CIE No.17.4 (*ILV: International Lighting Vocabulary*. 1987).

2.2.2.3 Resolution

Resolution is a spatial image attribute that is concerned with the fine detail reproduction ability of an imaging system. In analogue imaging, the most common measure of resolution is resolving power. It is measured using various test charts containing line (or bar) pairs with different line widths and is expressed in line pairs per mm (lp/mm).

In digital imaging, the term ‘resolution’ is also used as a descriptor of system performance, but in a slightly different way. For example, in a digital image, resolution describes the number of picture elements (or pixels) the image possesses. It is often expressed as the dimension of image horizontally and vertically, e.g. 200 pixels × 300 pixels. For digital capturing and display devices, it is also expressed by the number of

pixels per picture (or per unit distance), e.g. 5 megapixels, 100 dpi (or ppi), along with the physical dimension of sensor or displayable area. When physical dimension is provided, the resolution does not only provide information on available picture elements, but also information on the ‘fineness’ of the image, i.e. how small these pixels are.

Caution has to be taken when the term ‘resolution’ is used in digital imaging. Strictly speaking, the system performance depends not only on the ‘digital’ resolution, but also on the resolving power of the optical systems. The combined true resolution is often referred to as the effective system resolution (Wang and Hardeberg, 2012, Pierson *et al.*, 1996).

Resolution correlates with sharpness in general; however it is not the only image attribute which affects the definition of detail. Resolution is strongly dependent upon contrast, noise, as well as aspect ratio of the test target, along with exposure, processing, and observation condition (Ford, 1997, p.19, Heynacher and Kober, 1976, Jenkin, 2011c, p.434).

2.2.2.4 Sharpness

Sharpness is a spatial image attribute that is concerned with the edge reproduction ability of an imaging system. Various objective measures are used for the evaluation of sharpness, as shown in Table 2-3 (in Section 2.2.2) (Chapter 7 of Dainty, 1974). However, the *Modulation Transfer Function* (MTF) is the most sophisticated measure, which describes the relative contrast reproduction with respect to spatial frequency (Ford, 1997). There are widely used measuring methods for MTF determination in imaging. The wave recording method, which uses charts containing a series of sine-waves or a series of square-waves at various frequencies with known input modulation,

is one of them. Another is the edge method, which employs a captured (slanted) edge. The edge method is based on the Fourier theory of image formation (Burns, 2000) and the fact that a ‘perfect edge’ contains an infinite number of frequencies and therefore it is a perfect input signal for testing the spatial frequency response of a system (Jenkin, 2011c, p.447).

In addition to above methods, a new method is currently being developed using a target called “dead leaves” (Burns, 2011, Cao *et al.*, 2009). This new method is based on the theory that the MTF of the system can be derived from the system’s measured noise power spectrum (Dainty, 1974, p.255-258); it is designed to measure the *Spatial Frequency Response* (SFR) of digital cameras and mobile phone cameras based on the measured image texture (McElvain *et al.*, 2010).

Sharpness is discussed more in depth in Section 2.4, and is researched extensively in Chapter 5 with respect to displayed image size.

2.2.2.5 Noise and digital artefacts

Image noise is unwanted (random) fluctuation of light intensity. In chemical photography, these are due to the random structures, or clusters of silver grains in photographic materials (Jenkin, 2011c, p.435). In digital imaging, there are various sources of noise. Noise in digital images may be introduced in imaging hardware, by image processing, or can be part of the signal itself. Nakamura (Nakamura, 2006) summarised the causes of various noises in sensor stage, as presented in Table 2-5.

Temporal and fixed pattern noise are two common digital noise caused by the imaging sensors. Temporal noise is a random variation of reproduced signals that

fluctuate over time. It is often seen on images captured especially with higher ISO settings, regardless of the shutter speed (Cambridge in colour, 2013a).

On the other hand, fixed pattern noise (FPN) appears at certain pixel positions. In addition to the FPN caused by defective pixels in sensors, it is caused by non-uniformity of dark current over the whole pixel arrays, and/or by variations of performance of active transistors in imaging sensors (Nakamura, 2006, p.68). It is often seen on images captured at long exposures or high temperature (Nakamura, 2006).

	Dark	Illuminated	
		Below saturation	Above saturation
Fixed pattern noise (FPN)	Dark signal non-uniformity	Photo response non-uniformity	
	Pixel random	Pixel random	
	Shading	Shading	
	Dark current non-uniformity (Pixel-wise FPN) (Row-wise FPN) (Column-wise FPN)		
Temporal noise	Defects		
	Dark current shot noise	Photon shot noise	
	Read noise (Noise floor) Amplifier noise, etc. (Reset noise)		
			Smear, Blooming
	Image lag		

Table 2-5. Noise in image sensors, adapted from Nakamura (Nakamura, 2006, p.67).

In addition to image noise, digital artefacts are also considered as digital image noise. Triantaphillidou *et al.* (Triantaphillidou *et al.*, 2007) have identified a list of artefacts, their causes, and susceptible image areas. These are presented in Table 2-6.

Image Artefact	Cause of artefact	Susceptible image areas
Contouring	Poor quantisation	Uniform areas, slow varying areas (flat areas)
Jaggedness / Pixelisation	Insufficient spatial resolution	Slanted edges, slanted lines, high frequency information
Aliasing	Sampling	Areas with periodic high frequency information (high frequency lines)
Blocking	Discrete cosine transform (DCT) compression	Areas with high frequency information (busy areas)
Smudging / Colour bleeding	Discrete wavelet transform (DWT) compression	Areas with high frequency information (busy areas)
Ringing or edge echoes	Digital sharpening or DCT compression	Edges, lines
Patterning	Dithering	All areas expecting pure black and pure white
Streaking	Pixel to pixel non-uniformity in linear arrays (mostly of digital writing devices)	Uniform areas, slow varying areas (flat areas)
Banding	Cyclical variations in a property of digital writing devices	Uniform areas, slow varying areas (flat areas)
Colour misregistration	Optical images for different colour channels not geometrically identical	Small amounts: edges, lines, areas with high frequency information
		Large amount: all areas
Flare	Stay light in dark areas	Dark areas surrounded by high intensity areas

Table 2-6 Common digital image artefacts, their sources, and areas within images which are more susceptible to those artefacts. Note: Susceptible areas are defined here as either those affected mostly by the artifact or areas in which the artifact is more evident, adapted from Triantaphillidou *et al.* (Triantaphillidou *et al.*, 2007).

2.2.3 Image quality metrics (IQMs)

The objective image quality measures mentioned in the previous sections are based on the evaluation of individual perceptual image attributes. Even though these measures may correlate with the overall perception of image quality in general, they do not fully quantify, or predict it.

Image quality metrics (IQMs), on the other hand, are objective measures designed to produce a single value (or a set of values) aiming to describe or predict the overall image quality of images and systems (Triantaphillidou, 2011a, p.361-363). They may combine several physical measures derived from images or/and imaging systems (as, in Table 2-3), with attributes of the HVS (Granger and Cupery, 1972). The idea has been illustrated by Jacobson and Triantaphillidou (Triantaphillidou, 2011a) and shown in Figure 2-5.

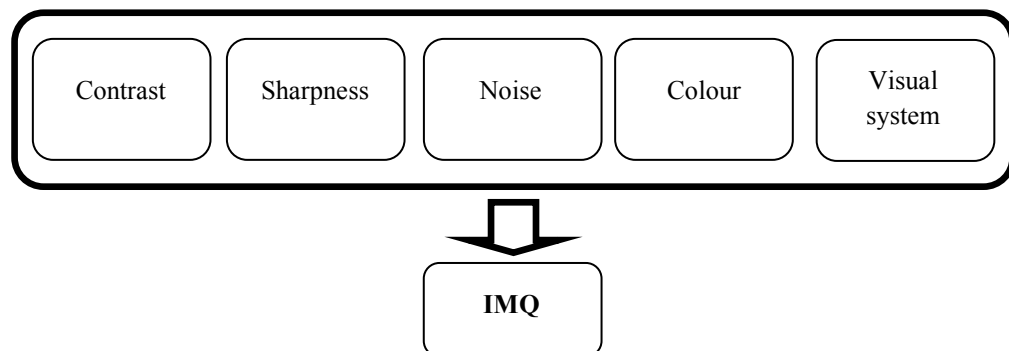


Figure 2-5. Measures or models describing the images' or the imaging systems' attributes and models of the HVS are used in IQMs, adapted from Triantaphillidou (Triantaphillidou, 2011a, p.361).

Hundreds of IQMs, metrics, and models have been proposed over the last 50 years. They differ in the numbers and types of physical measures that they use and the way in which they are combined with parameters of the human eye (Jacobson and

Triantaphillidou, 2002). Classical metrics, first designed and implemented successfully in analogue imaging, include the Subjective Quality Factor (SQF) by Granger and Cupery (Granger and Cupery, 1972), and the Square-Root Integral (SQRI) by Barten (Barten, 1990). They were based on the MTF of the imaging system and the spatial properties of the HVS. The SQRI_n by Barten (Barten, 1991), a modified model of the author's older version SQRI, and the Perceived Information Capacity (PIC) by Töpfer and Jacobson (Töpfer and Jacobson, 1993) take into account the combined system signal-to-image noise ratio. Both PIC and SQRI_n have been implemented with some success in the prediction of digital image quality (Ford, 1997, p.131-139). More recently, Jenkin *et al.* (Jenkin *et al.*, 2007) published the Effective Pictorial Information Capacity (EPIC). This metric is based on the spatial frequency responses of the chain that also include the HVS and noise.

There are also various metrics for colour image quality evaluations. The Colour Reproduction Index (CRI), proposed by Pointer and Hunt (Pointer and Hunt, 1994), is based on differences in hue (ΔH), lightness (ΔL), and chroma (ΔC), by taking the viewing conditions into account. In the late 1990's, the S-CIELAB, a spatial extension of the CIELAB, was proposed by Zhang and Wandell (Zhang and Wandell, 1996). Since it includes a spatial blurring stage using a pattern-colour separable method (and using the contrast sensitivity function of the HVS) prior to the evaluation of reproduction error, the measure corresponds better to the perception by the human eye (Zhang *et al.*, 1997).

A completely different approach to design metrics has been proposed by Bovik and his team at the University of Texas (Wang *et al.*, 2003, Li and Bovik, 2009, Sheikh and Bovik, 2006). It is based on the processing of image information rather than the

quantification of the imaging system properties. These types of metrics are based on assumptions made about the statistics and structural content of natural scenes, as well as the ability of the HVS to extract and interpret such structural information. These metrics quantify visual distortion between an original scene (or image) and a reproduction, and when there is a lack of original they used a statistical representation of the scene. Strictly speaking, they are fidelity measures, but are often referred to as quality metrics in the literature. Examples include the Structural Similarity Index (SSIM) (Wang *et al.*, 2003, Wang *et al.*, 2004) and the Visual Information Fidelity (VIF) (Sheikh and Bovik, 2006, Sheikh *et al.*, 2005, Sheikh and Bovik, 2005).

2.3 Subjective evaluation

In this section, an overview of psychophysics is given, along with scaling techniques and related experimental methods. Factors affecting image appearance are also identified and described.

2.3.1 Overview of psychophysics and psychometric scaling

Quantification of image quality in the past was focused more on objective image evaluation that was based on physical measurements (Engeldrum, 2000, p.5). It was based on the hypothesis that the results obtained by these evaluations well correlated with perceived image quality. Subjective image quality measures are based on the visual impression of image quality. They are a function of the HVS and the quality criteria of the observer (Triantaphillidou, 2001, p.32). Visual psychophysics is used to evaluate objective image quality.

Psychophysics deals with the measurement of the human response to physical stimuli (Johnson and Fairchild, 2002, p.124). Lockhead (Lockhead, 1992) defines psychophysical scaling models as having the form $R=f(I)$, with R the response and I the intensity of a (physical) attribute. The field of psychophysics has a long history. However, it caused considerable controversy from physical scientists in the nineteenth century (Boring, 1942). A century later, Fechner provided his ideas concerning the subjective measurement and the methodology (Fechner, 1966).

Since, it has been proven that valuable and accurate results can be obtained by implementing appropriate psychophysical methods; psychophysical investigations are commonly employed using various psychometric scaling techniques.

2.3.2 Scale types

Several types of measurement scales can be obtained from various scaling techniques. Four common types of subjective scales, with operational, structural, and statistical ascriptions were developed by Stevens (Stevens, 1946). In a classical textbook on image psychophysics by Engeldrum (Engeldrum, 2000), the author provided a summary of the types of scale, the operations and transformations related to these scales. They are presented in Table 2-7.

The nominal scale is obtained solely by categorisation with numbers, names, or labels, even though it is not much use for the purpose of subjective quality quantification. Care has to be taken, especially when numbers are used for such scales, i.e. phone number, sport players' number, etc., as they are quantitatively meaningless. Images identified by subject, such as portraits, landscapes, cityscapes, etc. can be

considered as ‘nominally scaled’. The nominal scale is highly useful for labelling purposes, although the scale does not possess any mathematical or arithmetic properties.

Scale type	Operations	Permissible Transformations
Nominal	Determination of equality	$y=f(x)$, any one-to-one transformation
Ordinal	Determination of greater or less than	$y=g(x)$, any monotonic transformation
Interval	Determination of equality of intervals or differences (distance)	$y=ax+b$, any linear transformation
Ratio	Determination of the equality of ratio	$y=ax$, any constant scale factor

Table 2-7. Stevens’ classification of scale types, adapted from Engeldrum (Engeldrum, 2000, p.45).

The ordinal scale is used to place items in an ascending, or descending order. It is a useful scale, but with limitations, in that the order omits any meaningful distances along the scale. It is obtained by rank order along some variable and thus has a ‘greater than’, or ‘less than’ property (Engeldrum, 2000, p.46). Unlike interval or ratio scales, ordinal data derived from images can tell us that a version of an image is considered of a better or worse quality than another version, but not ‘how much’ better or worse it is.

The interval scale is an ordinal scale possessing the property of distance (i.e. it possesses equally spaced intervals). The differences in sample scale values represent perceptual differences between two sample images, with respect to one perceptual attribute, or the overall image quality. Therefore, it is capable of specifying the equality of differences having the same visual significance (Triantaphillidou *et al.*, 2007, p.38). However, interval scales are floating scales that provide relative scale values (Triantaphillidou, 2011a, p.355), thus they must be distinguished from the ratio scales, which have a fixed point.

Finally, the ratio scale is an interval scale with the additive constant, or origin, often equal to zero (Triantaphillidou, 2011a). Unlike interval scales, these scales do not float with respect to the scale's origin. Engeldrum (Engeldrum, 2000, p.47) indicated that a zero point may not be experimentally measurable at all times (e.g. hue or image quality). Stevens (Stevens, 1946) has illustrated the psychometric scales as shown in Figure 2-6.

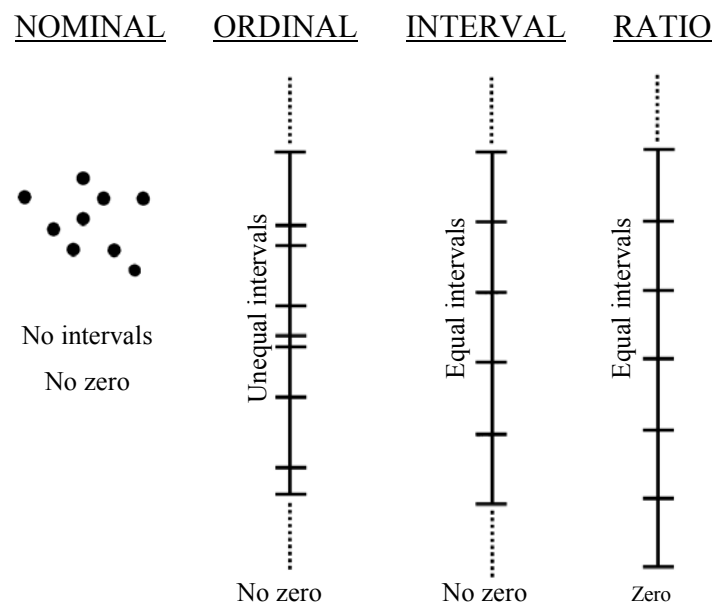


Figure 2-6. Illustration of psychometric scales, adapted from Stevens (Stevens, 1946).

2.3.3 Scaling methods

There are various known methods to obtain psychometric scales by subjective evaluations. The choice of methods depends on the purposes and time spent for the evaluation, since the implementation of different methods may have some disadvantages over others. Different scaling methods may even produce different results, although theoretically this should not be the case when experiments are well designed

and analysis is meticulously conducted (Boynton, 1961). Methods are described for two different purposes: 1) threshold and the JND evaluation (i.e. evaluation of fidelity) and 2) supra-threshold evaluation (evaluation of quality). Brief descriptions of common methods are presented in this section.

2.3.3.1 Threshold evaluation

There are two types of threshold evaluations: those concerned with absolute threshold and those concerned with just noticeable differences (JNDs). The absolute threshold evaluation is used to evaluate the detectability in ‘ness’ (i.e. how much of the stimuli is needed to just produce the ‘ness’ sensation). The JND evaluation is employed to measure the actual ‘ness’ differences that are observed.

Various scaling methods can be used to determine thresholds and JNDs. Classical methods are the *method of limits* and the *method of adjustments*. These were originally defined by Fechner (Fechner, 1966). The method of limits can be employed for image stimuli that are spaced closely with increased or decreased ‘ness.’ Observers are asked to answer whether they can detect the differences between stimuli possessing such an attribute. The main merit of this method is its efficiency, because the result can be obtained by just a few observations. Similar to the method described above, the method of adjustments can be used, where observers are asked to report on the visual differences by adjustment of the ‘ness.’ Observers may be asked to match the ‘ness’, or modify it until they can detect the differences, by using adjustment tools, such as a slider.

Another popular scaling method is the *paired-comparison method*, in which a pair of test stimuli are displayed one at a time and the observers need to reply with a

‘yes’, or a ‘no’ depending on their detection of differences. Although it produces very reliable results, it is rather impractical for a large number of stimuli.

Engeldrum (Engeldrum, 2000) has illustrated a typical psychometric curve with some critical points and regions, shown in Figure 2-7, as a function obtained from threshold experiments that describes the visual response to increasing ‘ness’ (c.f. Section 2.2.1). The absolute threshold and the point of subjective equality is taken where the proportion of the ‘yes’ responses of the observers are 50%. The range of proportion between 25% and 75% are described as the interval of uncertainty. The range of proportion between 50% and 75% are described as the just noticeable difference (JND). In this research, the JND was taken where the proportion of the ‘yes’ response was 75%, which is common in image quality related implementations (Keelan, 2002, p.50, Engeldrum, 2000, p.60).

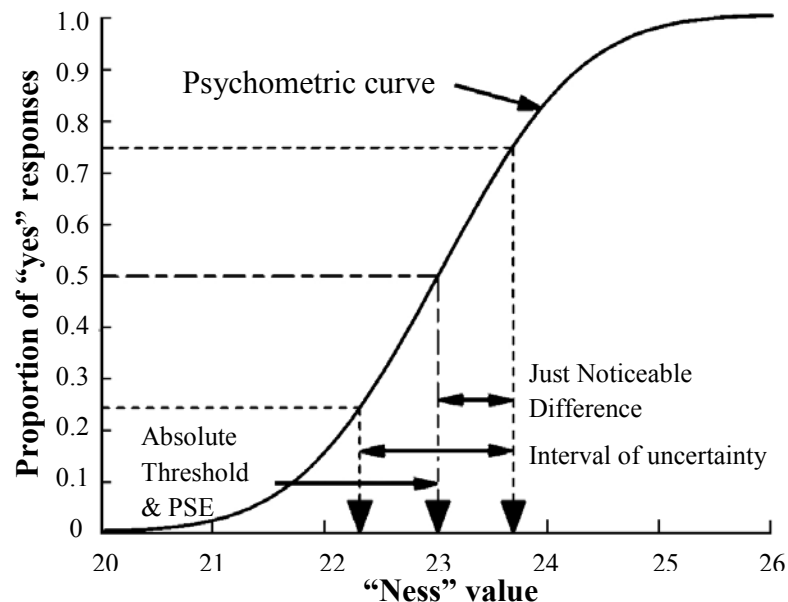


Figure 2-7. A typical psychometric curve. The axis is the value of a ‘ness’ and the ordinate is the proportion or probability of observers responding “yes”, adapted from Engeldrum (Engeldrum, 2000, p.56).

2.3.3.2 Supra-threshold evaluation

Supra-threshold evaluation methods are concerned with the development of subjective (or psychometric) scales. Different types of scales are obtained by different scaling methods.

The *rank-order method* is one of the simplest methods for obtaining an ordinal scale. The observers are asked to simply rank the image samples according to the degrees of one perceptual attribute (or ‘ness’) the image samples possess, or the overall image quality of the samples. Although the results obtained directly using this method will only contain the sequence of the image samples regardless of the significance of differences, there are various methods to transform the ranking data to interval scales by statistical analysis (Engel drum, 2000, p.109). Caution need to be taken when choosing this method, as it is rather impractical for a large number of image stimuli or attributes. Also, due to the physical dimension of display devices, the use of such method was limited to hard copy images in the past. This method is simple to implement and practical for a relatively small number of image stimuli, or attributes. It was used in this project to rank the impact that changes in display image size had on six image attributes. What were essentially ranked were the attributes, not the images, whilst one image was displayed at a time.

The *paired-comparison method* is another method of obtaining an ordinal scale. It is based on the law of comparative judgements (Triantaphillidou, 2011a), which relates the outcome of paired-comparison experiment to the perceptual differences between the stimuli and the uncertainty of perception, without reference to the physical origin of the differences (Engel drum, 2000, p.5). The observers are asked to simply

select one of two samples presented to them according to the degree of one perceptual image quality attribute the image samples possess, or the overall quality of the stimuli. This method is time-consuming when the number of test images is large. The number of the observations is rapidly increasing with the number of samples (i.e. $n(n-1)/2$ pairs for n number of samples (Boynton, 1984, p.359)). Therefore, this method is rather impractical for hardcopy samples and is mainly used for softcopy images. Ordinal scales, obtained from rank order experiments, can be turned into interval or ratio scales by using further statistical analysis.

The *category scaling method* is one of the simplest, easiest, and quickest methods to obtain scales. It is based on the *law of categorical judgement* (Thurstone, 1927), which relates the relative position of test stimuli to a number of categories. Image stimuli are viewed one at a time and observers are asked to place them in one of several categories (e.g. 'high', 'good', and 'bad' quality) or in categories denoted by numbers (e.g. quality 1 to 5, 5 being the highest). An ordinal scale and an interval scale can be obtained by category scaling. Interval scales obtained using such scaling is based on various assumptions, such as that category distances are perceptually equal. However, this is rarely the case in reality (Triantaphillidou, 2011a). In order to turn categorical scaling data to true interval scale, further statistical analysis may be needed for defining category boundaries between categories.

Lastly, the *magnitude estimation method* is a method of obtaining a scale by the estimation of quantities the test images possess (Stevens, 1946, Stevens, 1951). A reference image may be displayed at the beginning of the assessment. Then, observers are asked to assess the test samples, one at a time with respect to the reference image by using numbers. An interval scale or a ratio scale can be created using this method.

However, observer calibration by normalisation of the resulting data may be required, prior to the comparison of data. This is due to the variation in response by the individual observers (Engeldrum, 2000, p.139).

2.3.4 Visual matching technique

Visual matching is another common technique based on psychophysics. It is a simple, yet a powerful method for visual evaluation. It is used for the evaluation of changes in image quality or image appearance. It is strictly designed for image fidelity measurements. However, the method has been adapted and employed in image quality evaluation as well. The *method of adjustment* and the *paired-comparison method*, described in Section 2.3.3.1, can be employed for visual matching. A reference image and one or more test images are viewed at a time. The observers are asked to match the appearance or quality of the test images to that of the reference image. The International Organization for Standardization Technical Committee 42 (ISO-TC42) has approved and published recommended viewing conditions (*Photography--Psychophysical experimental methods for estimating image quality--Part 1: Overview of psychophysical elements*. 2005) and a number of visual matching techniques to evaluate the image quality (*Photography--Psychophysical experimental methods for estimating image quality--Part 2: Triplet comparison method*. 2005, *Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005).

2.4 Measuring and modifying sharpness

As discussed earlier, sharpness is concerned with the edge reproduction ability of an imaging system.

In conventional photography, *Modulation Transfer Function* (MTF) and *Acutance* are two common measures in sharpness evaluation. Acutance is used to evaluate sharpness of the light sensitive materials (Stroebel and Zakia, 1993, p.5). It is measured by evaluation of the mean square density gradient, divided by the density across an edge which is obtained from a microdensitometer trace. The degree of image sharpness depends on the shape and extent of the edge profile (Axford, 1988, p.345). MTF is a function taking into account the reduction in modulation (or image contrast) with respect to spatial frequency. There are several methods available for MTF measurement based on sinewave and edge targets as well as image noise test target (c.f. Section 2.2.2.4). The determination of MTF, however, is dependent upon the selected method (Triantaphillidou *et al.*, 1999).

In digital imaging, the Spatial Frequency Response (SFR) measure is widely used despite the fact that it is a measure strictly valid for linear systems (Burns, 2000, Burns and Williams, 2002). Traditionally, the MTF is obtained by imaging a ‘perfect’ edge and when this is not done, the MTF is ‘corrected’ for the original edge target frequency (Dainty, 1974, p.241). In the measurement of the SFR using slanted edge, modulus values are obtained by discrete Fourier transform of line spread function which is derived from edge profiled from the image data. Normalised modulus from the above step is the measured MTF (Williams and Burns, 2014).

Optical imaging systems are linear and isotropic. However, digital sensors and image signal processing (ISP) are often non-linear, non-stationary, and anisotropic (Yoshida, 2006, Triantaphillidou *et al.*, 1999). Corrections for various system non-linearities are implemented, when possible, for more accurate SFR determination. To compensate for system non-linearities of the capturing device, linearisation of the digital

image data using the Opto-Electronic Conversion Function (OECF) (*Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)*. 1999) is necessary. OECF is discussed in Section 2.5.1. The SFR evaluation method used in this work, along with the image sharpness adjustment methods are described in the following section.

2.4.1 SFR evaluation

Due to the physical nature of digital sensors, as discussed earlier, the techniques commonly used in conventional imaging are difficult to implement in digital imaging. For example, the edge of the test target has to be perfectly aligned with the pixel array which is a very difficult task. Due to the difficulties of implementing the traditional techniques, the slanted edge technique was developed for sampled systems.

In this research, the slanted edge method is used for the evaluation of system SFR. The technique is based on the traditional edge technique designed by Reichenbach *et al.* in 1991(Reichenbach *et al.*, 1991) for the determination of the *Spatial Frequency Response* (SFR) of digital capturing systems. The ISO first adapted this technique in 1999 and it was revised in 2000 (*Photography--Electronic still picture cameras--Resolution measurements*. 1999, *Photography--Electronic still picture cameras--Resolution measurements*. 2000).

Nowadays, SFR is used widely, since it provides various useful measures for system design and analysis (Jenkin, 2011c, Estriebeau and Magnan, 2004, Koren, 2006, Bang *et al.*, 2008). Also, the slanted edge based SFR measurement is adapted in various standards for the measurement of digital scanners and printers (*Photography--Electronic scanners for photographic images--Part 2: Film scanners*. 2004, *Information*

technology--Office equipment--Measurement of image quality attributes for hardcopy output--Binary monochrome text and graphic images. 2001, Information technology--Office equipment--Test charts and methods for measuring monochrome printer resolution. 2001).

A flowchart of the ISO standard implementation, which allows the automatic deviation of the SFR using a software application and a captured low contrast slanted edge, is illustrated in Figure 2-8.

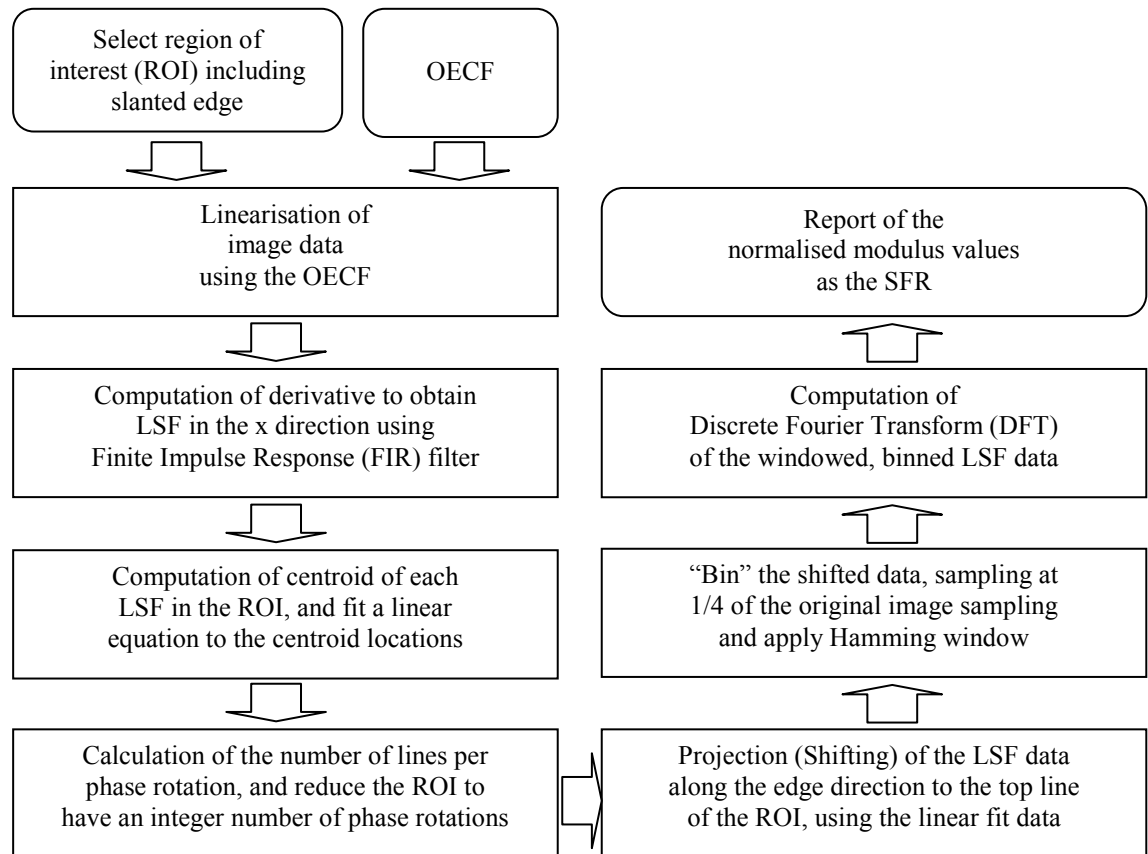


Figure 2-8. Flowchart of the deviation of digital SFRs from captured slanted edges, adapted from ISO 12233:2000 (*Photography--Electronic still picture cameras--Resolution measurements. 2000*).

2.4.2 Image sharpness manipulation

Image sharpness manipulation refers to the processing of increasing or decreasing image sharpness. In digital imaging, sharpness manipulation can be done in both spatial and frequency domains. In this section, commonly used filtering techniques for both sharpening and blurring are discussed. Also, a method that was employed in this project to produce sets of images with desired JNDs in perceived sharpness and blurriness, referred to here as *softcopy ruler images*, is described.

2.4.2.1 Filtering in spatial domain

Filtering in the spatial domain refers to the processing of images on the image plain (i.e. spatially). Filtering techniques in the spatial domain are categorised into linear and non-linear.

Linear filtering is based on the *discrete convolution* of a sub-image with the image. The sub-image is in the form of square matrix, consisting of digital values, as presented in Figure 2-9. It is referred to as ‘filter’, ‘mark’, or ‘kernel’, and its digital values are referred to as ‘*coefficients*’. The effect and the magnitude of the filtering depend on the filter coefficients. Pixel values on the original image are replaced by new pixel values, using reversible neighbourhood processing (Allen, 2011). More sophisticated filtering can be achieved by combining different techniques. However, these processes may not be reversible.

Unlike linear filtering techniques, non-linear filtering techniques are often non-reversible. One of the most commonly used non-linear filters is the median filter. It is mainly used for noise reduction by replacing the original pixel values by the median value of the neighbourhood pixels with a small cost of the loss of edge details.

$1/9 \times$	1	1	1
	1	1/2	1
	1	1	1
$1/16 \times$	1	2	1
	2	4	2
	1	2	1
	0	1	0
	1	-4	1
	0	1	0
	1	1	1
	1	-8	1
	1	1	1

Figure 2-9. Examples of commonly used spatial domain linear filters. Blur filters (left top, left bottom) and Laplacian sharpening filters (right top, right bottom), adapted from Gonzalez and Woods (Gonzalez and Woods, 2002, p.120-129).

2.4.2.2 Filtering in frequency domain

Filtering in frequency domain requires the Fourier transformation of image. It is based on the convolution theorem, which states that “the Fourier transforms of a convolution of two functions is the product of the Fourier transforms of these two functions” (Jenkin, 2011a). Jenkin has illustrated the convolution and the spatial frequency equivalent, shown in Figure 2-10.

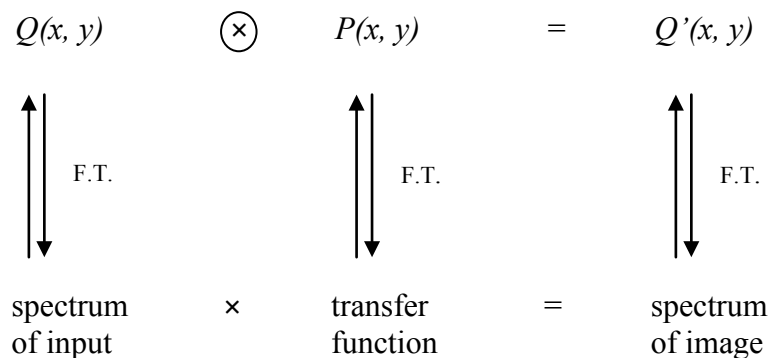


Figure 2-10. The imaging equation (convolution) and the spatial frequency equivalent, adapted from Jenkin (Jenkin, 2011b, p.133).

$Q(x, y)$ represents an image, $P(x, y)$ a spatial domain filter, and $Q'(x, y)$ represents a filtered image. Processing in spatial domain involves time consuming operations, whilst filtering in the frequency domain requires a single multiplication of the image spectrum with the filter spectrum, in addition to the forward and inverse Fourier transformations. The number of operations required to perform convolution in spatial and in frequency domains have been presented in Figure 2-11.

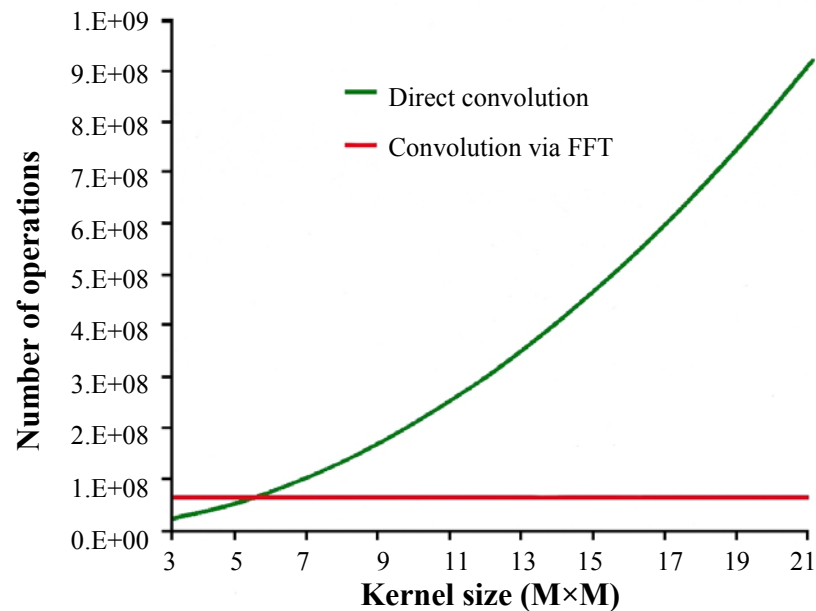


Figure 2-11. Number of operations required to perform convolution in spatial and frequency domains on a 1024×1024 pixel image versus kernel size, adapted from Jenkin (Jenkin, 2011a, p.524).

Typical lowpass (blurring), and highpass (sharpening) Fourier filters are presented with their images in Figure 2-12.

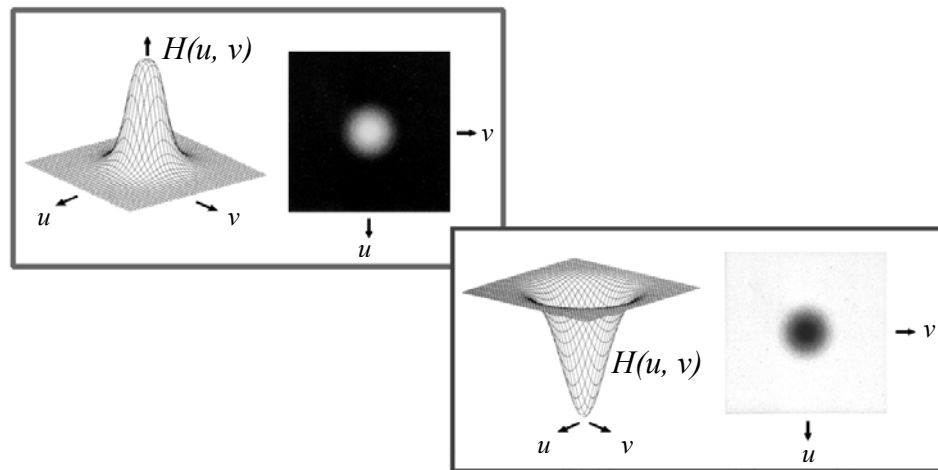


Figure 2-12. Perspective plots of a Butterworth lowpass filter (left), and a Gaussian highpass filter (right) transfer functions with their images, adapted from Gonzalez and Woods (Gonzalez and Woods, 2002, p.176-181).

2.4.2.3 Softcopy ruler method for generating sharpened and blurred images with known MTF

The theory of *softcopy ruler method* has its origin on work by Keelan (Keelan, 2000). The method was approved by the ISO, which produced three standards for the quantification of image quality (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005). ISO 20462-3 describes a methodology to create an image quality ruler, based on the performance of the imaging systems involved. The quality ruler comprises of a series of images with quantitatively known quality, in a single perceptual attribute. The quality ruler images are spaced by a constant JND interval in controlled viewing conditions. That means that, when the ruler varies in sharpness, or noisiness, the viewing distance must be fixed.

If the selected image attribute is sharpness, the quality of the ruler images is quantified by both the horizontal and vertical MTFs of the complete imaging system. Detailed procedures to generate ruler images are described below.

The system MTF conforms closely to the monochromatic MTF of an on-axis diffraction-limited lens, $m(v)$, which is given by Equation 2.16,

$$m(v) = \frac{2}{\pi} \left(\cos^{-1}(kv) - kv \sqrt{1 - (kv)^2} \right) \quad \text{when } kv \leq 1 \quad (2.16)$$

$$m(v) = 0 \quad \text{when } kv > 1$$

where v is spatial frequency in cycles per visual degree (CPD) and k is a constant.

The constant, k , has a range between 0.01 and 0.26. A series of model curves can be plotted by varying k . Then the combined imaging system MTF is compared with the modelled curves to find a closest shape curve. Once k is determined, a relative quality JND value associated with the k constant can be obtained using Equation 2.17. Plots of curves generated by Equations 2.16 and 2.17 are illustrated in Figure 2-13.

$$JNDs = \frac{17249 + 203792k - 114950k^2 - 3571075k^3}{578 - 1304k + 357372k^2} \quad (2.17)$$

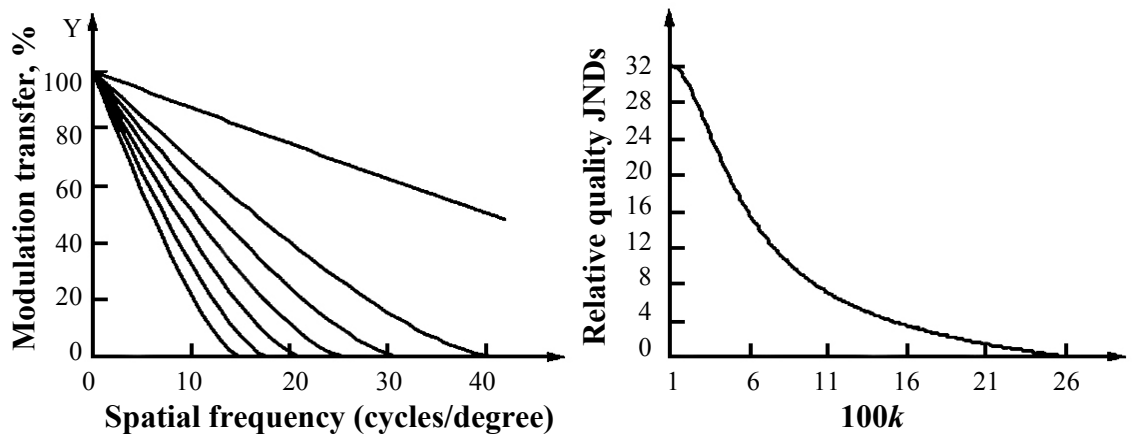


Figure 2-13. Plot of Equation 2.16, spaced by 3 JNDs (left) and Equation 2.17 (right), adapted from ISO 20462:3 (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005, p.10-11).

Once the constant, k , is determined, the quality of an imaging system can be quantified. Based on the relative quality JND found, a series of constants, k , with an interval of constant relative quality JNDs can be determined. The above steps are illustrated in Figure 2-14. In Figure 2-14, (a) represents the horizontal and vertical MTFs of the complete imaging system (in cycles/pixel), (b) represents the average MTF of the combined imaging system (in cycles/degree), (c) is a series of MTFs with varying k values for a diffraction-limited lens system based on Equation 2.11, and (d) is a plot of the relative quality JNDs versus the constant k , (e) represents a series of blurring filter functions, spaced at a constant interval, as described in ISO 20462-3. Vertically flipped versions of these filters are used to generate sharpened images. Graphical illustrations of the filter functions are presented in Figure 2-12.

The filter functions are in the form of an exponential function. The filter operation in a frequency domain is described in Equation 2.18,

$$H = 1 - e^{(-a \times b^D)} \quad \text{for blurring} \quad (2.18)$$

$$H = 1 + e^{(-a \times b^D)} \quad \text{for sharpening}$$

where D is the digital image size of the image's spectrum, and a , and b are the variables representing the sizes of the filter apertures.

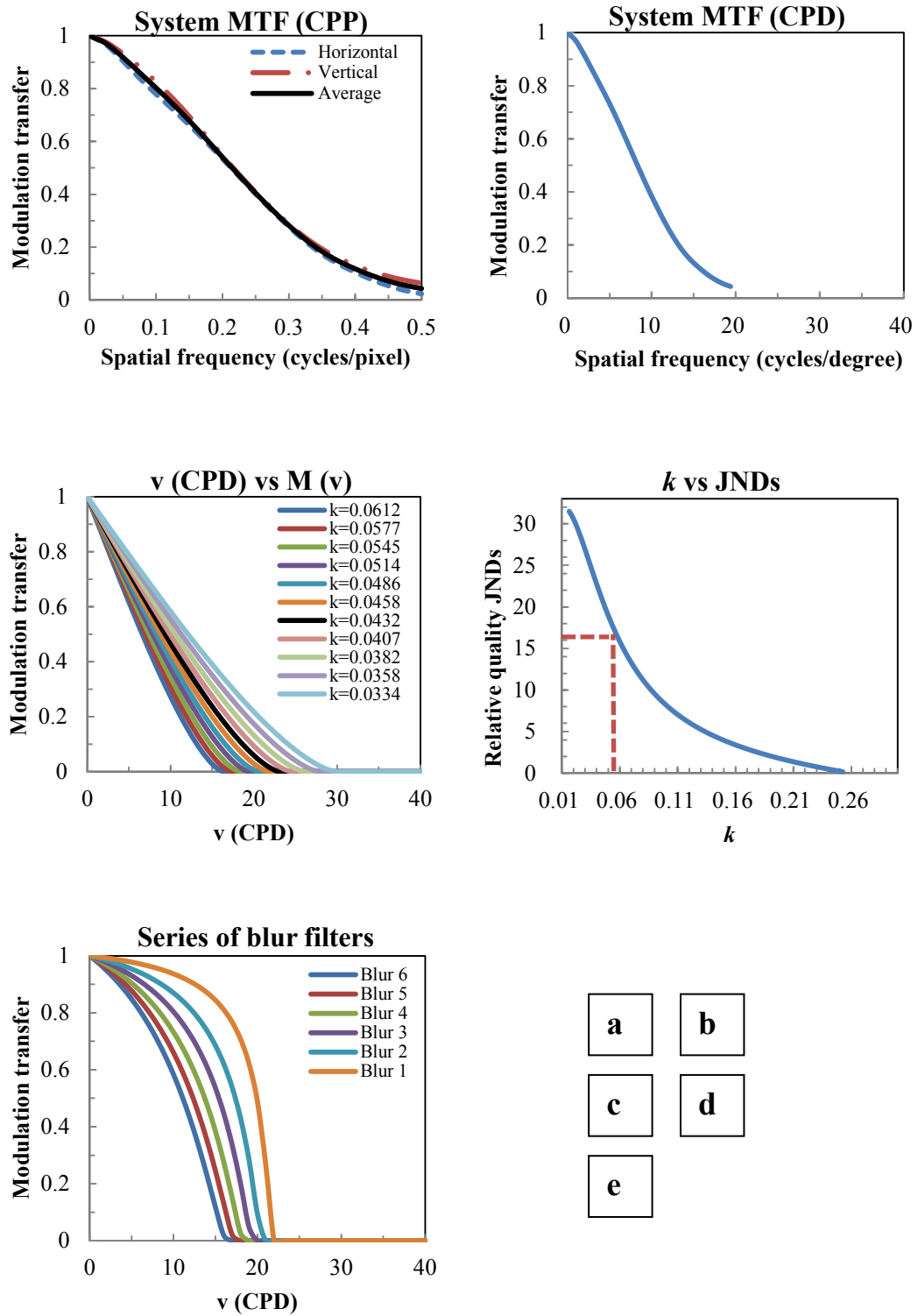


Figure 2-14. Implementation of steps involved in the creation of frequency domain Gaussian filters, with a constant interval based on ISO 20462-3 (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2012).

2.5 Measuring and modifying tone reproduction and contrast

In digital imaging, various definitions and evaluation metrics for contrast are available. However, most of the definitions assign a single value to describe the contrast of the whole image, regardless of the fact that the contrast may vary across the image (i.e. local contrast versus global contrast) (Peli, 1990). In addition, metrics and formulae for the evaluation of image contrast should take into account visual contrast perception. The study of various approaches to link physical contrast with visual contrast perception is ongoing (Peli, 1990, Triantaphillidou *et al.*, 2013).

In this section, the methods used for the evaluation of the tone reproduction of imaging systems (c.f. Chapter 3) and the concept of gamma (the descriptor of system contrast, c.f. Section 2.2.2.1) in tone reproduction is briefly explained. Common formulae used in contrast evaluation; along with contrast enhancement techniques are also described.

2.5.1 Opto-Electronic Conversion Function (OECF)

The opto-electronic conversion function is used to describe the relationship between the input luminance and the output pixel value in capturing devices (*Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)*. 1999). Although the native photo-electronic conversion characteristics of digital sensor materials exhibit approximately linear response to the light intensity, most digital cameras have non-linear characteristics (Yamada, 2006, p.118, Cheung *et al.*, 2004). This is often imposed by manufacturers in form of near inverse relationship to the non-linearity of CRT display devices (Westland *et al.*, 2012, p.144), which is commonly adopted by LCD devices. Also to use the available bit-depth

more efficiently, that is in accordance with the HVS response to luminance (Poynton, 1996, p.113).

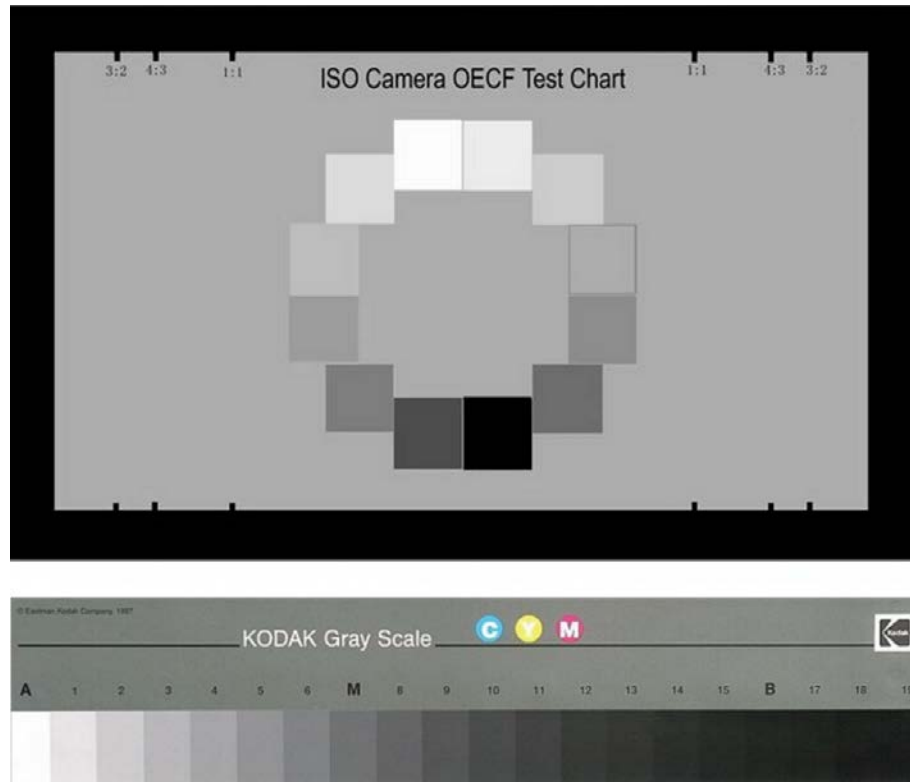


Figure 2-15. Two different test charts for measuring transfer functions of acquisition devices. Top: ISO camera OECF test chart. Bottom: Kodak Q-13 greyscale, adapted from Triantaphillidou (Triantaphillidou, 2011b, p.385).

ISO first published a standard method for measuring the OECFs in 1999 (*Photography--Electronic still picture cameras--Resolution measurements*. 2000). The OECF of digital camera systems can be measured by capturing an ISO OECF test chart, or a Kodak Q-13 chart, both presented in Figure 2-15. These charts consist of a number of uniform grey patches of various density levels. The relationship is often plotted in log-log units, i.e. log pixel values vs. log luminance, or in linear units, i.e. pixel values vs. target reflectance.

Since the measurement of Spatial Frequency Response (SFR) is strictly applicable to linear systems, as described earlier (c.f. Section 2.4), the gamma value derived from the OECF is used in the SFR implementation to compensate for the non-linearity of the camera system.

2.5.2 Electro-Optical Transfer Function (EOTF)

The electro-optical transfer function is used to describe the relationship between the input voltage and the output luminance in displays. Although the native electro-optical transfer characteristics of LCDs exhibit an S-shape form, similar to the photographic characteristic curve, as shown in Figure 2-16, most of LCD devices produce a response, which is imposed by manufacturers in hardware, or software to mimic the characteristics of CRT displays (Fairchild and Wyble, 1998, Chapter 2 of Bala, 2002, Day *et al.*, 2004).

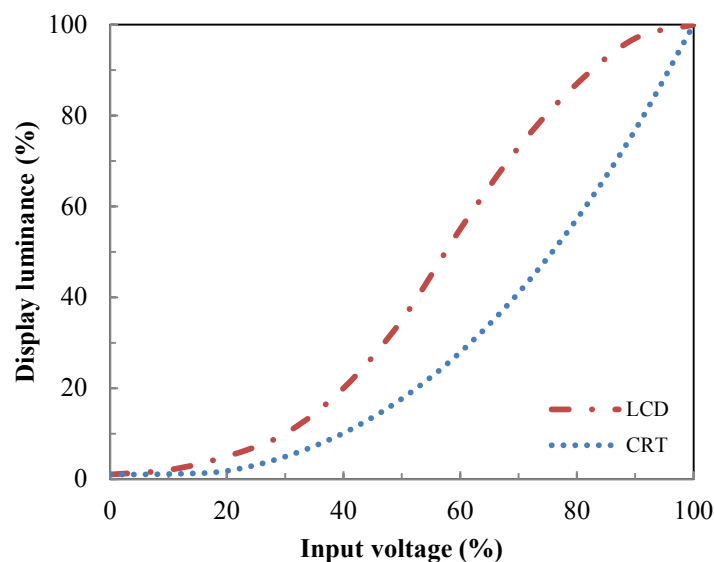


Figure 2-16. Typical electro-optical transfer functions for CRT and LCD devices, adapted from Glasser (Glasser, 1997).

As described in earlier section for CRTs (c.f. Section 2.2.2.1), the EOTF is also plotted in linear units and described by a power function, in which the exponent represents gamma, γ .

2.5.3 Formulae for contrast evaluation

In this section, three commonly used formulae to evaluate contrast are explained.

2.5.3.1 Michelson contrast

Michelson contrast (C_M) (Michelson, 1962), is used for measuring the physical contrast of a simple pattern with bright and dark features such as single frequency sinusoidal gratings (Peli, 1990, Triantaphillidou, 2011b). C_M is measured using Equation 2.19,

$$C_M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (2.19)$$

where L_{max} and L_{min} are the highest and lowest luminance in grating, respectively.

2.5.3.2 Weber fraction definition of contrast

Weber fraction (definition of contrast) is used for measuring local contrast of an area with uniform luminance on a uniform background. C_W is measured using Equation 2.20,

$$C_W = \frac{\Delta L}{L} \quad (2.20)$$

where ΔL is the increment (or decrement) in the target luminance from luminance of uniform background, L .

2.5.3.3 Root mean square (RMS) contrast

Root mean square (RMS) contrast is a calculation of standard deviation of luminance values. It is often normalised by the mean image luminance to return values between zero and one. Unlike the Michelson formula and the Weber fraction of contrast, RMS contrast can be used to define the contrast of compound grating images and of complex digital images (Tiippana *et al.*, 1994, Moulden *et al.*, 1990). The RMS contrast is known to be a good predictor of the relative apparent contrast (Triantaphillidou *et al.*, 2013, Bex and Makous, 2002).

$$C_{RMS} = \sqrt{\left[\frac{1}{R \times C} \sum_{x=1}^{C-1} \sum_{y=1}^{R-1} (I_{xy} - \bar{I})^2 \right]} \quad (2.21)$$

where R and C are the number of rows and columns in the image, I_{xy} is the normalised luminance of x^{th} y^{th} pixel, \bar{I} is the mean normalised luminance of the image.

RMS contrast (C_{RMS}) of a two dimensional image can be defined by the root mean square deviation of the pixel luminance from the mean pixel luminance of the image, divided by the image dimension (Pavel *et al.*, 1987). C_{RMS} of a two dimensional image can be calculated using Equation 2.21, adapted from Peli (Peli, 1990).

2.5.4 Image contrast enhancement

Image contrast enhancement refers to the process of increasing or decreasing image contrast. Similar to sharpness enhancement, this can be done in both spatial and frequency domains. Various enhancement techniques are available to enhance image contrast on a global and local scale. In this section, spatial domain contrast enhancement techniques are discussed.

2.5.4.1 Histogram equalisation

Histogram equalisation is a “technique that generates a grey map which changes the histogram of an image and redistributing entire pixel values to a user specified histogram” (Hassan and Akamatsu, 2004). It is based on the assumption that important information in the image is contained within areas of high probability of distribution (i.e. the Probability Density Function, PDF). Although, this is one of the most widely used techniques for contrast enhancement, the brightness of an image can be also changed when using it (Kim, 1997).

2.5.4.2 Contrast stretching by piecewise linear transformation function

Contrast stretching by piecewise linear transformation function is one of the simplest techniques to enhance image contrast. The technique is carried out with a set of linear functions which are characterised by the fact that the input-output function’s slope is altered linearly between defined control points. This technique is also used to enhance the contrast of one, or more subjects, which comprise grey levels in a certain range (Allen, 2011, p.504). The idea behind this technique is to increase the dynamic range of the grey levels in the image (Gonzalez and Woods, 2002, p.85). This is achieved by

applying functions with a gradient of lower than 1.0 to above and below the control points, and higher than 1.0 between the control points. The reverse effect can be obtained by using an inverse function to decrease contrast.

2.5.4.3 Contrast enhancement using an ‘S-shape’ function

Contrast enhancement using an S-shaped function is a similar process to contrast stretching by piecewise linear transformation functions. However, unlike piecewise linear transformation functions, an S-shaped function alters pixel values smoothly (Braun and Fairchild, 1999). Therefore, the processed images possess more natural tonal ranges across the entire tonal range.

2.6 Scene dependency and classification

Objective image measurements associated with image quality are based on the assumption that there is a fundamental relationship between these measurements and the subjective impression of one of the image attributes, or the overall image quality (Triantaphillidou *et al.*, 2007).

Objective quality measures however, do not always correlate to the subjective quality. In this section, scene dependency in subjective image quality is described, along with objective methods for scene analysis and classification, which can be used to compensate for scene dependency in objective quality models and metrics.

2.6.1 Scene dependency

As briefly discussed in Section 2.1, scene content is an important factor for evaluation of image quality. Triantaphillidou *et al.* (Triantaphillidou *et al.*, 2007) described three

types of scene dependency. The first type of scene dependency is due to the observer's preference (or quality criteria). Freiser *et al.* (Freiser and Biedermann, 1992) found that the sharpness is judged differently for portrait image and architecture scene. The second type of scene dependency is due to the visibility of noise, or other artefacts (c.f. Table 2-6) in some types of images (or image areas) compared with other images. Artefacts are more prominent on some images than on others. Therefore, in addition to the objective quantification of digital image noises or artefacts, their visibility of such artefacts should also be considered as additional image attribute (Keelan, 2002, p.131). The third type of scene dependency is due to variation in the output of digital process such as sharpening/blurring and image compression, which depends on the image content and scene features (Triantaphillidou, 2011a).

Scene dependency issues make it difficult to design psychophysical evaluations and analyse results for a variety of scenes. For this reason, many studies are conducted using the ISO set of test scenes (*Graphic Technology: Prepress digital data exchange--CMYK standard color image data (CMYK/SCID)*. 1997). However, this standard image set does not represent a wide range of images and a variety of scene content. A proposed method for overcoming scene dependency is scene classification with respect to image quality measurements. It has been widely researched in our laboratories with considerable success (Triantaphillidou *et al.*, 2007, Orfanidou *et al.*, 2008, Oh *et al.*, 2010).

2.6.2 Classification of scenes

Images and scenes can be classified into relatively small groups with respect to various scene characteristics that play a significant role in image quality measurements, e.g.

illumination characteristics, directional viewing aspect, spatial distribution of scene elements, and local illumination conditions, spatial frequency content, and colour content, etc. (Jones and Condit, 1941). One way to achieve this is by simply inspecting images and grouping them with respect to selected image attributes (Keelan, 2002). The other way is through objective image analysis. Triantaphillidou *et al.* (Triantaphillidou *et al.*, 2007) have conducted scenes analysis techniques that are directly relevant to image quality experiments, using various statistical measures and segmentation to classify scenes with respect to both spatial and colour attributes. Global and average intensities, global contrast, and ‘busyness’ of the scene were measured with respect to spatial attributes. In addition, the variance in chroma, VC_{ab}^* , was measured in CIELAB colour space and was proposed as measure of global image colourfulness, or global colour contrast. Several of the proposed measures have been found to correlate with the perception of image content (Triantaphillidou *et al.*, 2007, Orfanidou *et al.*, 2008, Mancusi *et al.*, 2010, Falkenstern *et al.*, 2011). Recently, Oh (Oh *et al.*, 2010) used second order statistics and edge analysis to classify scenes according to their susceptibility in noisiness and sharpness. He further used his objective classification to calibrate successfully a number of device-dependent image quality metrics to take into account scene dependency.

2.7 Appearance versus image size

Image appearance is a phenomenon of visual perception. Therefore, it is naturally affected by various factors including the surround viewing conditions for images as well as the physical changes of image size, or the changes in the angle subtending the observer’s eye. Many studies have been conducted to identify and quantify the changes

in image appearance with respect to the image size or viewing angle. As early as in the 1960's, Bartleson and Breneman (Bartleson and Breneman, 1967a) pointed out that change in image size affect the perceived contrast. Choi *et al.* (Choi *et al.*, 2007b, Choi *et al.*, 2007a) conducted psychophysical experiments using colour patches of various sizes under various illumination conditions, including dark, indoor, and also outdoor conditions. They confirmed that the colour appearance was affected by changing the patch size and the viewing conditions. Nezamabadi and Berns (Nezamabadi and Berns, 2006, Nezamabadi *et al.*, 2007) have investigated the effect of image size on the colour appearance of softcopy reproduction, using a contrast matching technique. They identified that lightness is mostly affected by the changes in image size and then in chroma. Xiao *et al.* (Xiao *et al.*, 2011) also confirmed that lightness and chroma are affected by changes in size. However, they found that size has no effect on hue appearance.

Similar results have been found in a recent study by Wang and Herdeberg (Wang and Hardeberg, 2012), who investigated the changes in appearance of all colour attributes as well as sharpness, noise and compression with changes in the visual angle.

The studies related to changes in appearance of pictorial images with changes in image size are nevertheless limited. The work presented in this thesis is dedicated to this subject and aims to provide answers to the following: questions of which image attributes are most affected when changing displayed image size; how perceived sharpness changes with altering displayed image; how perceived contrast changes with changing image size.

Chapter 3

Device characterisation

This chapter describes the characterisation and settings of devices used for image capture and image display. Device calibration is referred to the settings of an imaging device to a known state (Fairchild, 2005, p.316), for example to a chosen maximum luminance, white point, gamma setting, etc. Characterisation of an imaging device defines the relationship between input signals and the response of the device. Colorimetric characterisation of a device refers to the creation of a relationship between the device coordinates and a device independent colour space (Fairchild, 2005, p.316). For example, colorimetric camera characterisation defines the relationship between the camera response (in RGB) and the input tristimulus values. Similarly, colorimetric display characterisation defines the relationship between the resultant CIE measurement of display response and the input data (Johnson, 1996).

In this study, colour characterisation of both camera and display devices were carried out for the sRGB setting. sRGB is one of the most commonly available colour settings (default on most compact cameras, such as the Apple iPhone used in this work, and available in most DSLRs). The aim in this project was to produce and examine test-images for a colour setting employed by most consumer users. Secondly, for the camera, accurate colorimetric reproduction is not so important to this project, because the aim was not to produce colorimetric digital images but pleasant images, the appearance of which can be subsequently examined on display with respect to image size.

Tone reproduction and colorimetric characteristics were measured for the capturing devices. These measurements were based on ISO 14524 (*Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)*. 1999) and ISO 17321-1 (*Graphic technology and photography--Colour characterisation of digital still cameras (DSCs)--Part 1: Stimuli, metrology and test procedures*. 2006). In addition, the spatial frequency response (SFR) of the capturing devices was measured using the slanted edge technique described in ISO 12233 (*Photography--Electronic still picture cameras--Resolution measurements*. 2000).

For the characterisation of display devices, display characteristics suggested in BS EN 61966-4 (*Multimedia Systems and Equipment--Colour measurement and management--Part 4: Equipment using liquid crystal display panels*. 2000) were evaluated to determine the experimental methods and interface design. In addition, the positional non-uniformity at the observation plane was investigated in accordance with the psychophysical investigation set up. SFR measurements of the display devices are described in Chapter 5, along with their application in the development of the frequency domain filters for image sharpness enhancement.

3.1 Digital cameras

Two digital image capturing devices, one with an 8 megapixel sensor and another with a 2 megapixel sensor, exhibiting different overall image qualities, were used for recording a number of natural scenes. The purpose was to produce ‘identical’ image content for each scene with both cameras. The Canon EOS 30D digital SLR, equipped with an EF-S 10-22mm (35mm equivalent focal length of 16-35mm) lens allowed full access to camera function, such as aperture, shutter speed, ISO and custom white balance settings. It also allowed for saving the captured images in various file formats with, or without image compression. The Apple iPhone (1st generation) mobile phone camera was equipped with a fixed lens (35mm equivalent focal length of 35mm) with a fixed aperture $f2.8$. It had a default automatic white balance and did not allow access to the ISO, or shutter speed settings. It saved 24-bit sRGB images in JPEG format. Due to the limited access to the settings on the Apple iPhone camera, the characterisation of the Apple iPhone camera was carried out prior to that of the Canon camera; whilst the characterisation of the Canon camera was carried out for similar setup and settings to those of the Apple camera for consistency. The camera settings used for image capture are shown in Table 3-1.

	Canon EOS 30D	Apple iPhone
Pixel resolution	3504×2336 (8.2 MP)	1600×1200 (1.9MP)
Colour representation	sRGB, 24bits	sRGB, 24bits
ISO	100-1600	Information not available
Image format	JPEG	JPEG
Lens	EF-S 10-22mm at 22mm (FOV 63°)	Built-in lens (FOV 63°)
Aperture	$f4.5$ - $f11$	$f2.8$

Table 3-1. Camera settings for the image capture.

3.1.1 Tone characteristics (Opto-Electronic Conversion Function)

Tone reproduction characteristics were evaluated by measuring the opto-electronic conversion functions (OECF), using the methods described in ISO 14524. The standard describes two methods. One is the focal plane OECF method, which is used for camera with removable lenses. The main advantage of this method is that it provides an accurate measure of the OECF of imaging sensors and camera electronics under selected illumination conditions. An alternative OECF method is suggested for the cameras with non-removable lenses, for which exposures can be made using reflective test targets. Although the former method is recommended for accurate measurement, it was not suitable for characterising the camera with a fixed lens (Apple iPhone). Therefore, the alternative method was implemented for consistency.

A Kodak Q-13 greyscale test target, which contains 20 reflective neutral patches with approximately 0.1 density increments, was used for the purpose. The density of each of the patches was read using a calibrated Macbeth TR924 reflection densitometer; an average of 3 measurements from the central area of each patches were recorded. After the measurements, the test target was placed at approximately 150cm away from the sensor plane and occupied the central 4% of the frame. The immediate surroundings of the target were covered with neutral mid-tone background to minimise any unwanted colour effects caused by the background colour or flare. A pair of tungsten lamps was used for the standard copy lighting to illuminate the target evenly. The mean luminance of the target area was measured using a Minolta CL-200 chroma meter. The mean luminance of the target area was 2,485lux for both cameras.

As a preliminary step, a series of exposures was made to investigate the effects of ISO and aperture settings on tone reproduction by the Canon 30D camera. Colour

space was set to sRGB and colour temperature was set to automatic mode. The results showed that the effects of ISO and aperture settings were negligible, 0.0022% by the ISO and 0.0028% by the aperture settings. Therefore, the test target was captured using the same settings by both cameras under the studio set up described above.

Pixel values of the captured greyscale patches were measured and the \log_2 mean PVs were calculated. Density values were converted to luminance (L_i), cd/m², using Equation 3.1 (*Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)*. 1999).

$$L_i = \frac{10^{-D_i E}}{\pi} \quad (3.1)$$

where D_i is the grey scale patch visual density

E is the illuminance, in lux, incident on the chart

L_i is the luminance, in candelas per square metre, of the patch with density D_i

The equation is based on the assumption that the test target is a perfect reflector thus there is no loss in luminance. \log_2 PV was plotted against $\log_{10}L_i$ in Figure 3-1. The measured OECF showed slight variations for each of the channels, with larger standard errors in darker patches for both cameras. The linear portion (spanning the mid-tones, where original SFR target luminance values are falling) of the measured OECF had a gamma of $\gamma=0.592$ for the Canon 30D and a gamma of $\gamma=0.496$ for the Apple iPhone.

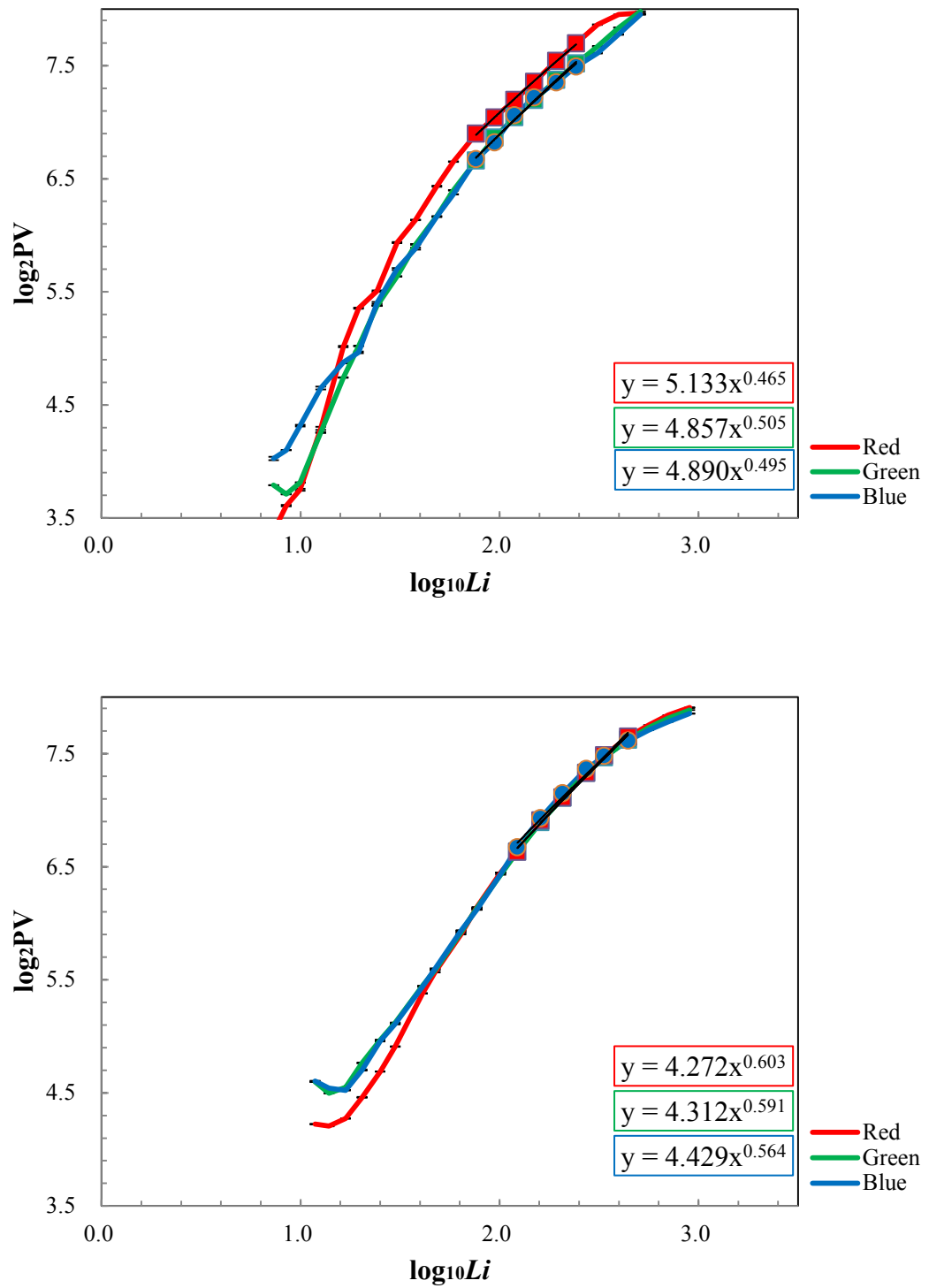


Figure 3-1. Tone reproduction characteristics of the Apple iPhone camera (top) and the Canon 30D camera (bottom). $\log_2 PV$ was plotted against $\log_{10} L_i$.

3.1.2 Colorimetric characteristics of sRGB output

The colorimetric characterisation of the capturing devices was carried out by adapting the conditions suggested in ISO 17321-1. The standard introduced two methods for the evaluation of colorimetric characteristics of digital still cameras. One is the spectral sensitivity-based method. This method requires quantified uniform illumination. The spectral (multi spectral) sensitivity-based method is suitable for accurate measurement of the response of imaging sensors when raw image data can be obtained (Cheung *et al.*, 2004, Berns and Shyu, 1995, Cheung and Westland, 2003). Alternatively, the target-based method can be used for which exposures can be made using reflective test targets with known spectral and colorimetric characteristics. This method is efficient and also suitable for both devices (Johnson, 1996, *Graphic technology and photography--Colour characterisation of digital still cameras (DSCs)--Part 1: Stimuli, metrology and test procedures.* 2006). Although the former method provides often more accurate measurement, it was not suitable for characterising one of the cameras with limited access to the settings. Therefore, the target-based method was implemented for consistency.

A GretagMacbeth ColorChecker Color Rendition Chart (McCamy *et al.*, 1976), which contains 6 achromatic patches with difference densities and 18 chromatic patches representing natural colours, was used for the purpose. Spectral reflectance and CIE 1931 XYZ tristimulus values were measured using the GretagMacbeth ColorEye 7000A spectrophotometer. The target was then photographed using both cameras, under similar photographing conditions to those employed for the tone characteristics as described in Section 3.1.1. The mean luminance was 1,833lux ($x=0.4609$ and $y=0.4128$, at a colour

temperature of 2700K). The white balance of both camera systems was set at automatic mode, since only automatic white balance was available on the Apple iPhone camera.

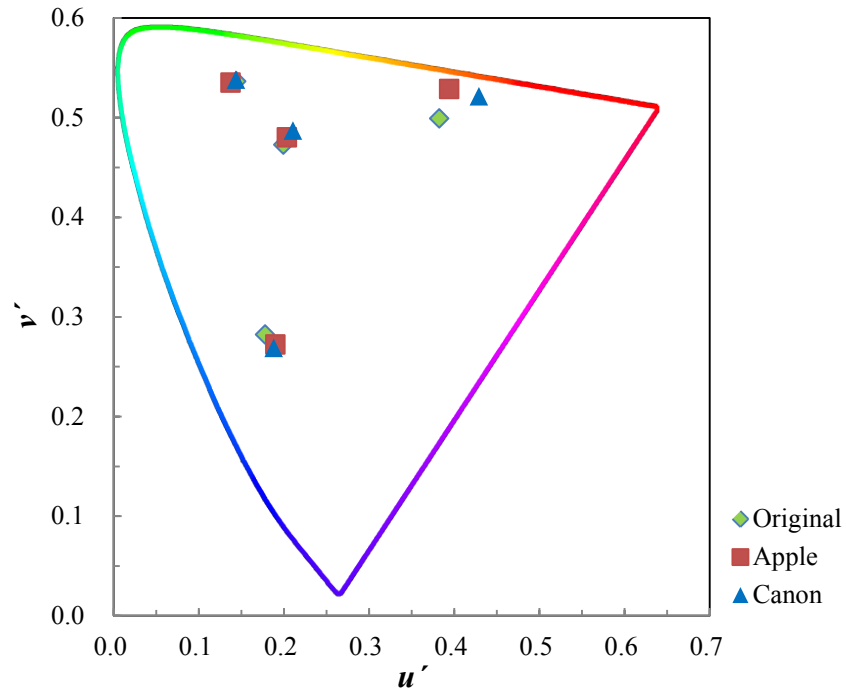


Figure 3-2. Original and the captured red, green, blue, and white patches of the GretagMacbeth ColorChecker by both cameras.

Average pixel values of the captured images were measured using NIH ImageJ image analysis software (Rasband, 2013). For the calculation of the colour reproduction errors of the capturing devices, mean pixel values of the captured patches in standard RGB were converted to XYZ tristimulus values by implementing Equations (2) to (7) from the BS ISO 61966-2-1 (*Multimedia Systems and Equipment--Colour measurement and management--Part 2.1: Default RGB colour space--sRGB*. 2000, p.10-12). CIELAB L^* , a^* , b^* coordinates were then calculated from the estimated the tristimulus values. Further, the CIE 1976 u' and v' chromaticity coordinates were calculated according to Equations 5.12 (a) and (b) (TriantaphillidouandAllen, 2011, p.89). Figure

3-2 illustrates the plotted colour coordinates of the captured primary colour patches and of the white patch, along with those of the original test target. It was clear that reproduction errors were larger for the red patch compared with the other full on primary colours.

Using the CIELAB coordinates, colour reproduction errors of both camera systems were calculated using the colour difference formulae. The commonly applied CIE 1976 ΔE_{ab}^* (cf. Equation 2.10) and the more perceptually uniform CIEDE2000 ΔE_{00}^* were used (Luo *et al.*, 2001).

The colorimetric performance of both cameras, when set to sRGB setting, was poor with ΔE_{ab}^* of over 3 for all patches. The errors were especially large for the saturated red and purple colours ($L^*=40-50$). The biggest differences were found on reddish patches, with ΔE_{ab}^* of 33.14 and 47.23 for the Canon 30D and the Apple iPhone, respectively. Also, the neutral patches appeared rather ‘reddish’, with colour differences ranging from 14.15 to 16.08 for the Canon 30D, and from 7.03 to 19.81 for the Apple iPhone. However, the colour patches containing ‘strong green and/or blue’ colours were captured with smaller colour reproduction errors. The colour differences for all colour patches were plotted in Figure 3-3, and the values are shown in Table 3-2.

	Canon 30D		Apple iPhone	
	ΔE_{ab}^*	ΔE_{00}^*	ΔE_{ab}^*	ΔE_{00}^*
Max	33.14	26.64	47.23	34.80
Min	5.23	2.10	3.31	2.22
Mean	15.42	10.34	18.48	11.41

Table 3-2. Colour differences between the original and captured patches for both camera systems.

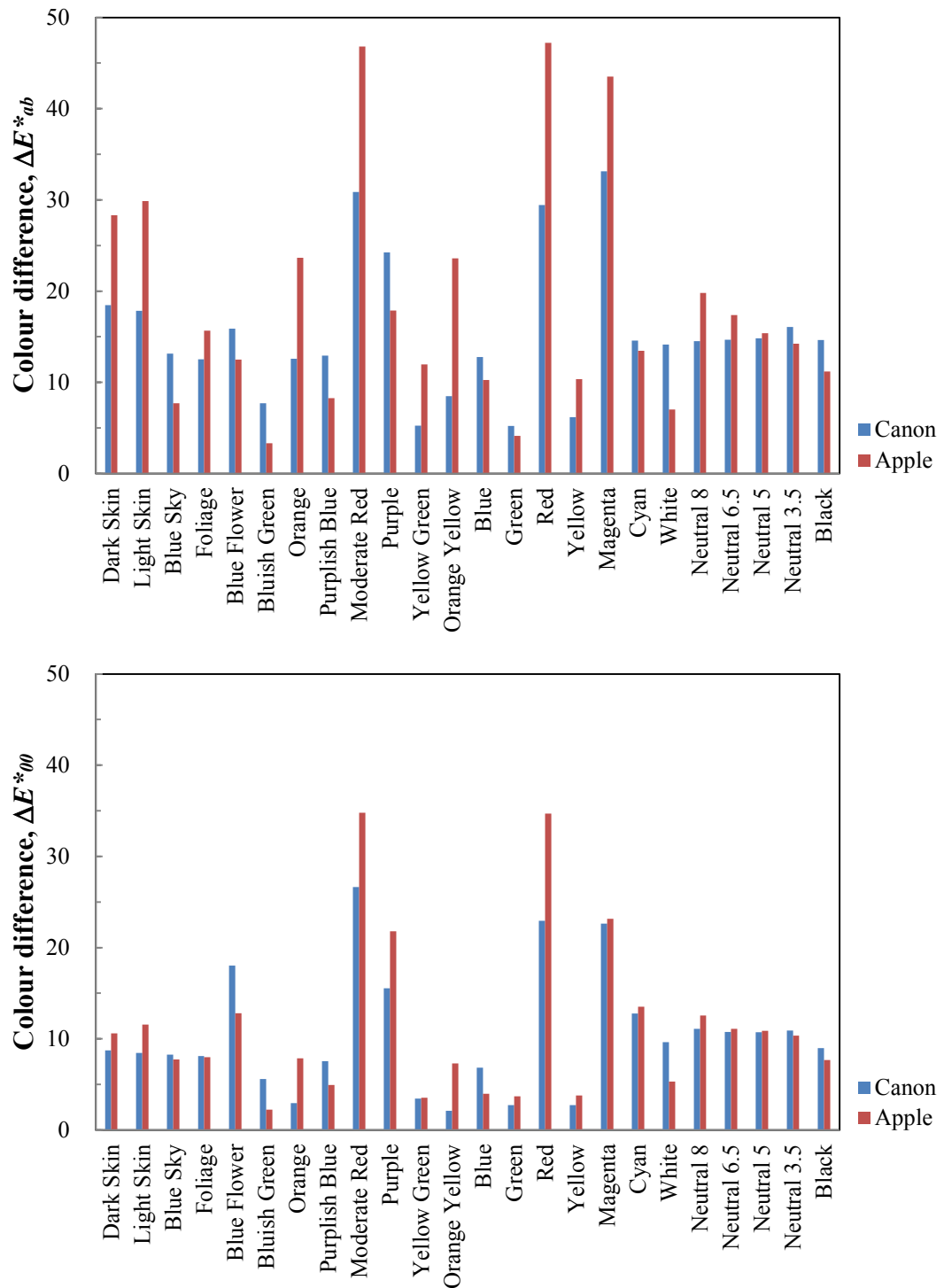


Figure 3-3. Colour production errors between the original and captured patches for both camera systems using two commonly used colour difference formulae.

3.1.3 SFR measurements using the slanted edge method

Spatial frequency responses (SFR) (*Photography--Electronic still picture cameras--Resolution measurements*. 2000) for both cameras, set at the selected settings, were measured using the slanted edge method. Although the standard originally recommended a high contrast test target (40-80:1), a simple test target containing a low contrast edges (contrast ratio of 3:1) was used (Burns and Williams, 2002, *Using SFRplus Part 1*. 2013). The target was fixed on a flat surface slanted approximately 5° at horizontal orientation then captured 3 times using both cameras under the standard copy lighting described in earlier sections. The capture was repeated for the vertical camera orientation. A Minolta CL-200 chroma meter was used to ensure the variation of the luminance across the target below $\pm 2\%$, as recommended in ISO 14524 (*Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)*. 1999).

A series of exposures was made with the Canon 30D at various aperture settings to investigate the effect of aperture on the SFR measurements. SFR was found to be high with the lens apertures up to $f11$ then it dropped dramatically at $f16$ and smaller (c.f. 5.1.2 Figure 5-2). SFRs calculated from the edges captured at $f5.6$ - $f11$ were found to be 7 JNDs in relative quality scale higher than those calculated from the edges captured at $f16$ - $f22$. Therefore, the aperture of the Canon 30D camera system was set at $f8$ to obtain the sharpest edge. The aperture of the Apple iPhone was set at $f2.8$ as a default.

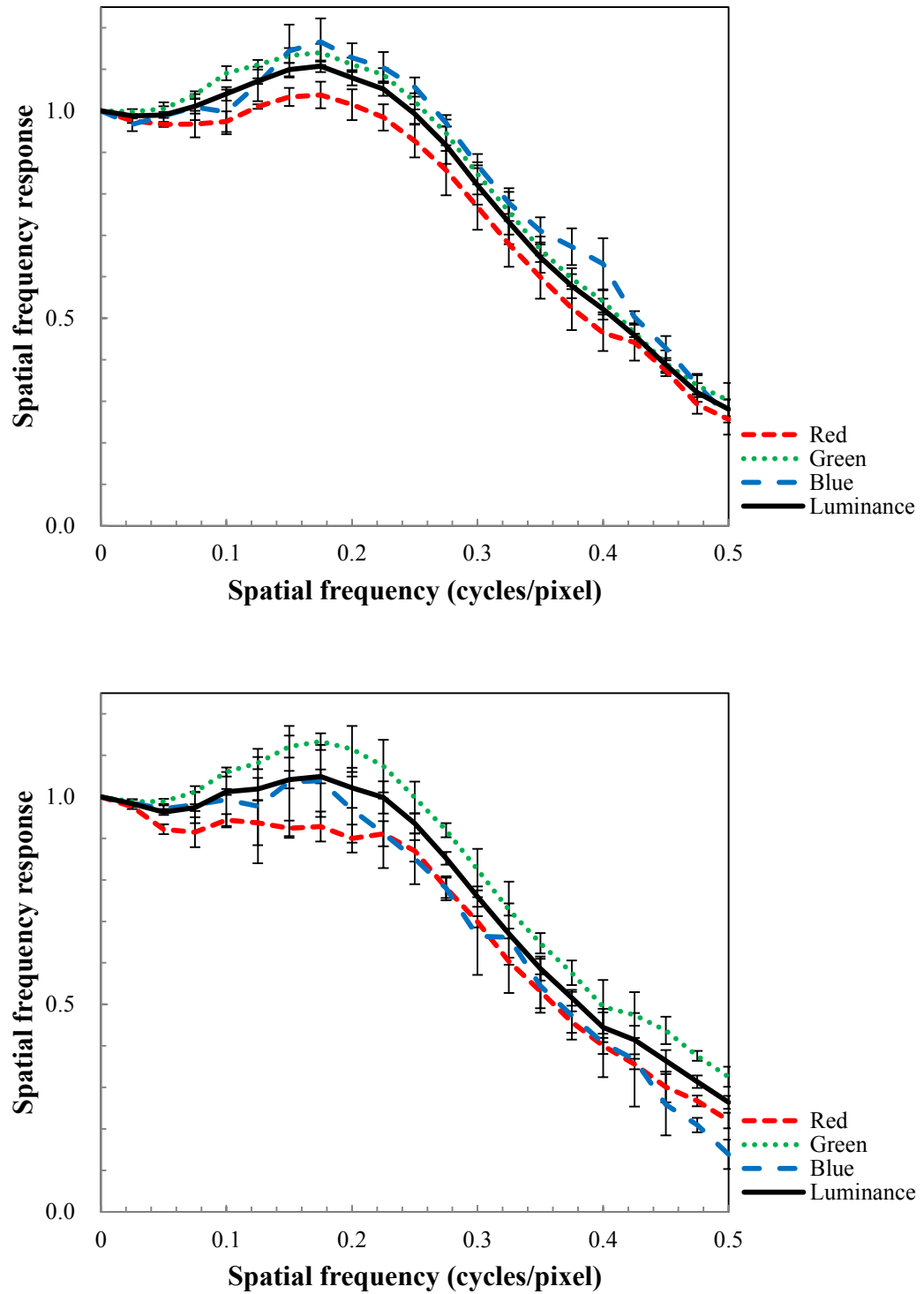


Figure 3-4. Horizontal (top) and vertical (bottom) SFR of the Apple iPhone camera.

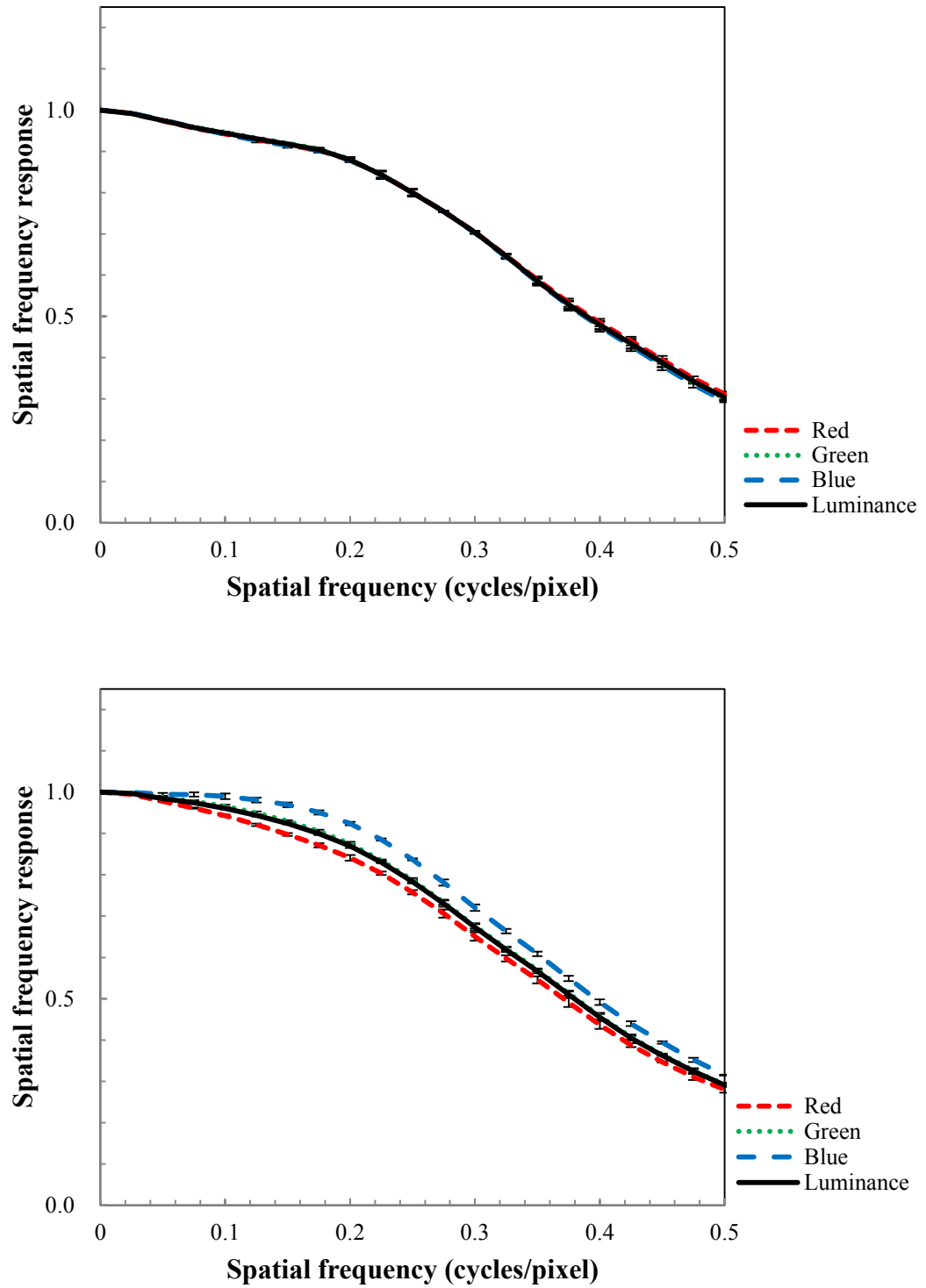


Figure 3-5. Horizontal (top) and vertical (bottom) SFR of the Canon 30D camera.

SFRs of captured edges were computed using Imatest image analysis software (*Imatest*. 2013) and the SFR up to the Nyquist frequency, 0.5cycles/pixel, were recorded. The camera system gamma, γ , derived in Section 3.1.1 was used as the gamma correction function for the data within the software (i.e. implemented to linearise the image data before the computation of the SFRs). For each camera, average SFRs were calculated for the horizontal and vertical orientations. The SFRs of the individual channel and in luminance are plotted in Figures 3-4 and 3-5.

The Canon 30D camera showed fairly even spatial frequency responses in all three channels at both orientations with small temporal variations. However, the SFRs of the Apple iPhone varied between colour channels at both orientations with much larger temporal variations. Also, the ‘hump’ at the low-mid frequencies was clearly caused by the edge enhancement. Overall, the SFR of the Canon 30D camera was found to be slightly higher than that of the Apple iPhone camera.

3.1.4 Summary

Two different format and overall quality cameras were characterised in terms of colour reproduction, tone reproduction and spatial frequency response. sRGB colour setting and JPEG format was chosen for both cameras (for consistency). This choice was made because these were the only setting and format available on the Apple iPhone camera. A pair of tungsten lamps was chosen as light source and using automatic white balance settings for the characterisations.

From the tone reproduction characterisation, the reproduction was also found rather poor for the dark patches (higher densities) but it was fairly even for the mid-tone and lighter patches (lower densities). The Apple iPhone gamma was found to be

$\gamma=0.496$, which with a typical sRGB display system gamma of 2.2, would give an overall system gamma of approximately $\gamma=1.09$. The Canon 30D system gamma was found to be $\gamma=0.592$, would give an overall system gamma of approximately $\gamma=1.30$. These meant that the overall contrast of the starting images in the following experiments was slightly different, with the images originating from the Canon 30D having higher contrast. This difference was not big enough to matter, since the images from each individual camera were tested in separate experiments.

Colour reproduction of both systems set to sRGB was found to be poor especially for the saturated red and the purple colours, with the maximum colour difference, ΔE_{ab}^* , of 33.14 and 47.23, for the Apple iPhone and the Canon 30D respectively. Colour reproduction errors for the greenish and bluish patches were much smaller, producing the minimum colour differences between the original and captured patches. The mean ΔE_{ab}^* of approximately 10 was found from both devices.

Although more accurate methods are suggested in ISO standards, alternative methods were adapted due to the limited access to the settings, and fixed lens feature on one of the cameras. Therefore, the results may have been varied under different lighting conditions, and using various settings.

However, the purpose of employing two camera systems exhibiting different image qualities was to produce pleasant images with identical image contents, rather than the colorimetrically accurate reproduction of natural scene. Even though the Canon 30D showed slightly better performance for the characteristics measured above, both systems were capable of producing pleasant images under various illumination conditions, which is what is required in the experiments.

3.2 Liquid crystal displays (LCDs)

Two In-Plane Switching (IPS) Liquid Crystal Displays (LCDs) from the same manufacturer were used in this study. An EIZO ColorEdge CG210 21.3'' LCD was used in research described in Chapter 4 and an EIZO ColorEdge CG245W 24.1'' LCD was used in work described in Chapters 5, 6 and 7.

The characterisation of both display devices were carried out under the environmental conditions described in BS EN 61966-4. The standard recommends the measurements of various important display characteristics that have the potential to influence psychophysical investigations such as these described in the following chapters.

3.2.1 Conditions of measurement, calibration and settings

The calibration of the display devices and all measurements were carried out with at least one hour warm up time as specified in BS EN 61966-4 except for the temporal stability characteristics which require measurement with a cool down period (*Multimedia Systems and Equipment--Colour measurement and management--Part 4: Equipment using liquid crystal display panels*. 2000, p.28-30). The room temperature was approximately 20° Celsius $\pm 3^\circ$, as measured in the beginning and in the end of measurements which was also used during the psychophysical investigations.

For the calibration and profiling of the displays, a GretagMacbeth Eye-One Pro was used. During the psychophysical investigations, the EIZO CG210 LCD was calibrated daily using the Eye-One Pro. A built-in calibration sensor was used for the daily calibration of the EIZO CG245W LCD. The technical specifications of the display

devices, the settings for the calibration, profiling and experiments are shown in Table 3-3.

	EIZO CG210	EIZO CG245W
Displayable area (cm)	43.2(H)×32.4(V)	51.8(H)×32.4(V)
Native pixel resolution (pixels)	1600(H)×1200(V)	1920(H)×1200(V)
Display colour	24bits from a palette of 30bits	24bits (DVI)/30bits (DP) from a palette of 48bits
Viewing angle (°)	170(H), 170(V)	178(H), 178(V)
Pixel pitch	0.27mm(H), 0.27mm(V)	
Maximum brightness	250cd/m ²	270cd/m ²
Maximum brightness for calibration and experiments	120cd/m ²	
Colour representation	sRGB	

Table 3-3. Technical specifications of display devices and the settings used during calibration and experiments.

A Konica-Minolta CS-200 tele-chroma meter designed especially for LCDs was used for the measurement of both displays. The chroma meter was connected to a PC, which drove the instrument using CS-S10w designated software (Konica-Minolta, 2013). The instrument was placed 150cm away from the centre of the display device, a distance little greater than that recommended in the standard (*Multimedia Systems and Equipment--Colour measurement and management--Part 4: Equipment using liquid crystal display panels*. 2000), in a plane parallel to that of the display. A set of 240 by 240 pixel patches with a different pixel value, as described in the standard, was created. The instrument was set to measure the luminance and the tristimulus values of displayed patches, with a 0.2° field of view, in slow mode. Three measurements were averaged

each time. Except for the positional non-uniformity characteristics, a small central area in horizontal and vertical orientations of the display was measured.

3.2.2 Tone characteristics (Electro-Optical Transfer Function)

The relationship between the output luminance and the input pixel values describe the tone reproduction of display devices. A total of 32 red, green, and blue patches with a pixel value interval of 8 were created. Also, a total of 32-step neutral ramp was created.

The XYZ tristimulus values for each patch were measured and then normalised. The normalised luminance output was plotted against the normalised input pixel values in linear-linear scale for individual channels (Red, Green, and Blue) and in combination (Neutral), as shown in Figures 3-6 and 3-7, for the EIZO CG210 and the EIZO CG245W respectively. The CG245W had excellent tone reproduction characteristics for each individual channels and also when all channels were combination with overall gamma of $\gamma=2.16$. However, the CG210 had inconsistent tone reproduction in each individual channels as well as when all channels were combined. The gamma of the CG210 was $\gamma=2.09$. High level of Z values was found for red and green channels on both displays, which is typically seen on LCD devices. Non-zero black levels on the LCDs are further discussed in a later section (c.f. Section 3.2.4).

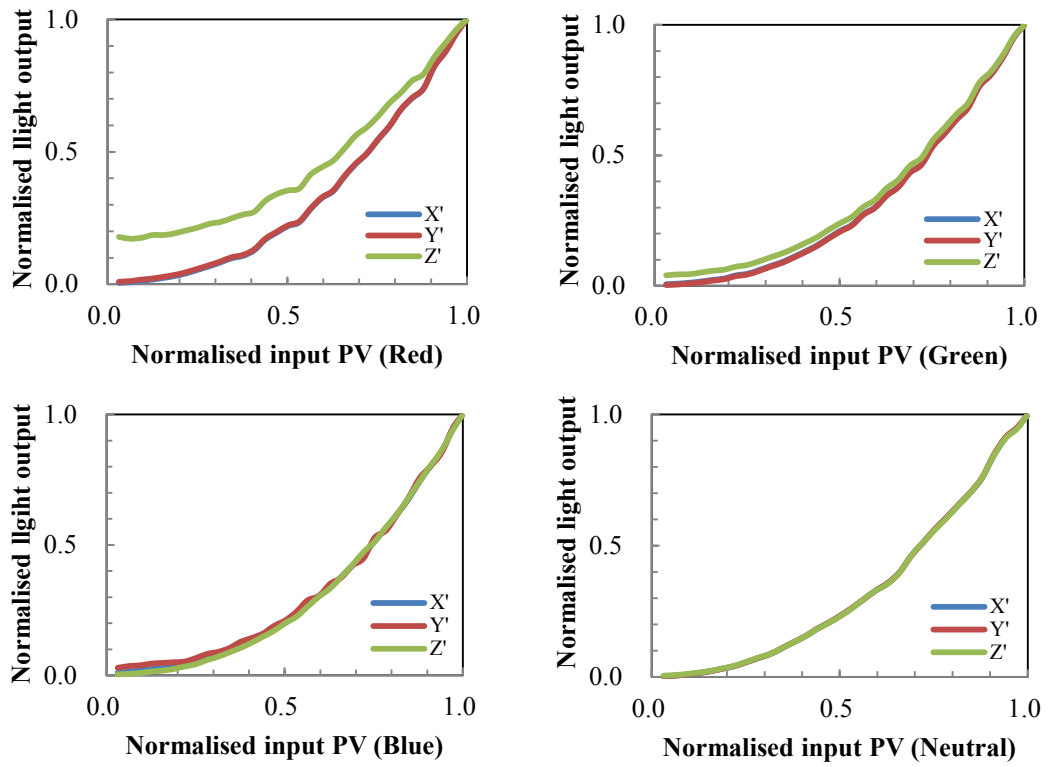


Figure 3-6. Tone characteristics of the EIZO CG210 display.

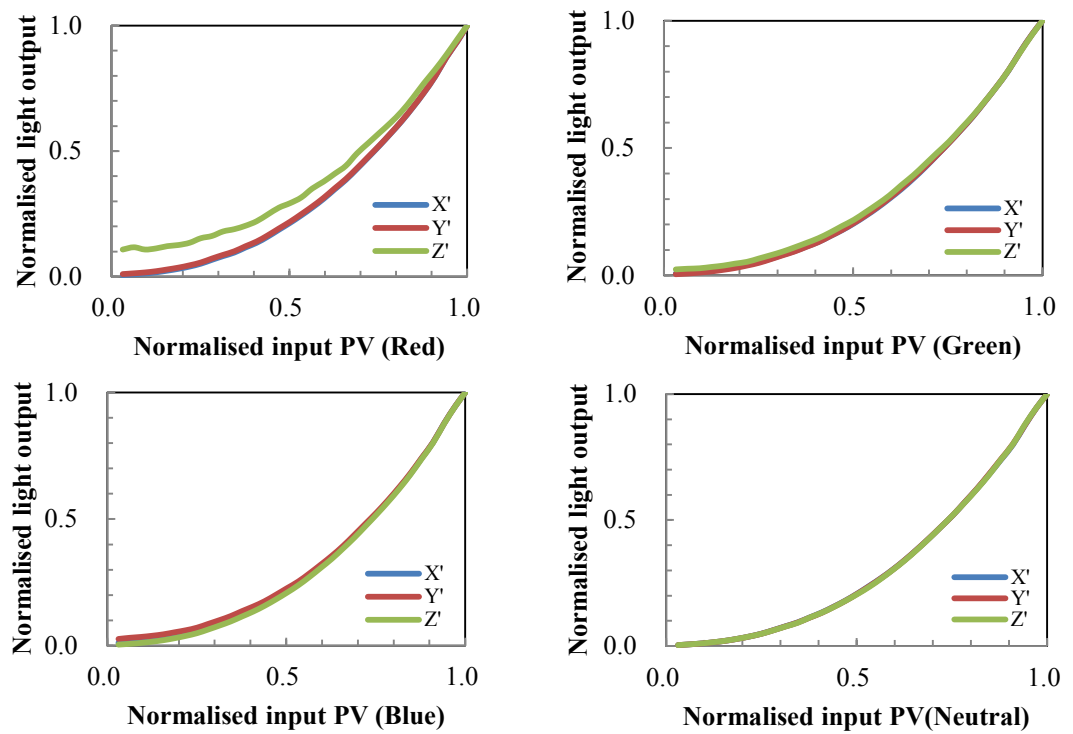


Figure 3-7. Tone characteristics of the EIZO CG245W display.

3.2.3 Basic colorimetric characteristics

Basic colorimetric characteristics of display devices describe the linear relationship between the output tristimulus values and the corresponding maximum input pixel values (c.f. BS EN 61966-4 Section 8).

A set of four patches, each containing full on primaries and pure primaries at full strength in the centre of the frame, was created and displayed on the calibrated display devices. The XYZ tristimulus values were measured. The measured tristimulus values were then normalised by the measured luminance value for the white, Y_n . CIE 1976 u' , v' coordinates of the reproduced patches were calculated from the measured tristimulus values and plotted in Figure 3-8. Corresponding peaks in sRGB colour space were also plotted for comparison purposes. The CIE 1931 tristimulus values and the CIE 1976 u' , v' coordinates are shown in Table 3-4.

	CG210					CG245W				
	X'	Y'	Z'	u'	v'	X'	Y'	Z'	u'	v'
Red	49.88	25.17	2.85	0.456	0.520	51.85	26.01	2.22	0.462	0.522
Green	37.63	78.76	12.16	0.120	0.565	41.50	85.19	11.45	0.127	0.566
Blue	19.89	7.53	109.22	0.178	0.147	22.46	8.73	117.89	0.177	0.155
White	105.59	109.8	123.01	0.199	0.466	114.18	118.54	131.04	0.177	0.436

Table 3-4. CIE 1931 tristimulus values and CIE 1976 chromaticity coordinates for the full on primaries and the white from both display devices.

Both display systems were calibrated to produce white luminance of 120cd/m² at D₆₅. The reproduction accuracy of all four colours on both systems was excellent when compared with the peak colours in sRGB colour space. The CG245W display performed slightly better than the CG210 display. The white point colour temperatures

were found at 6583K and 6455K on the CG210 and CG245W, respectively, although both systems were calibrated to 6504K.

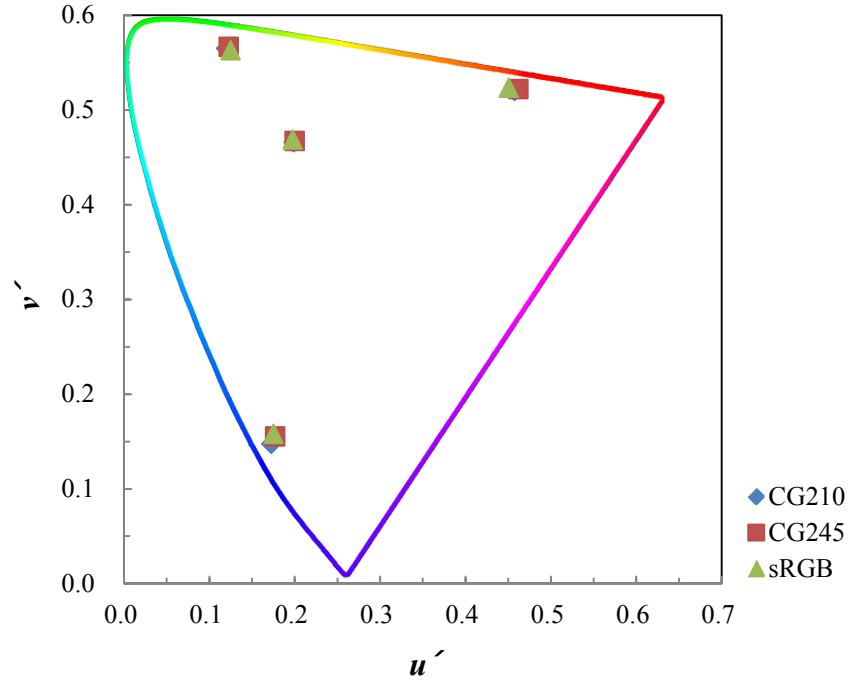


Figure 3-8. Reproduction of the full on primaries and the white on display devices and their corresponding values in sRGB colour space.

The values in Table 3-4 were used to derive elements of a 3×3 conversion matrix, S , defined as:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = S \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.2)$$

where R , G , B are the normalised input pixel values thus determined as;

$$S = \begin{bmatrix} x_R/y_R & x_G/y_G & x_B/y_B \\ 1 & 1 & 1 \\ z_R/y_R & z_G/y_G & z_B/y_B \end{bmatrix} \begin{bmatrix} S_R & 0 & 0 \\ 0 & S_G & 0 \\ 0 & 0 & S_B \end{bmatrix} \quad (3.3)$$

where S_R, S_G, S_B are the solution of Equation 3.4;

$$\begin{bmatrix} x_R/y_R & x_G/y_G & x_B/y_B \\ 1 & 1 & 1 \\ z_R/y_R & z_G/y_G & z_B/y_B \end{bmatrix} \begin{bmatrix} S_R \\ S_G \\ S_B \end{bmatrix} = \begin{bmatrix} x_W/y_W \\ 1 \\ z_W/y_W \end{bmatrix} \quad (3.4)$$

The derived coefficient matrix, S , for each display devices are shown in Equation 3.5a (CG210), 3.5b (CG245W), along with that of the sRGB (Equation 3.5c) (*Multimedia Systems and Equipment--Colour measurement and management--Part 2.1: Default RGB colour space--sRGB*. 2000).

$$S = \begin{bmatrix} 0.4461 & 0.3383 & 0.1796 \\ 0.2251 & 0.7080 & 0.0680 \\ 0.0255 & 0.1093 & 0.9860 \end{bmatrix} \quad (3.5a)$$

$$S = \begin{bmatrix} 0.4275 & 0.3469 & 0.1887 \\ 0.2144 & 0.7122 & 0.0734 \\ 0.0183 & 0.0957 & 0.9906 \end{bmatrix} \quad (3.5b)$$

$$S = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \quad (3.5c)$$

3.2.4 Colour tracking characteristics

Colour tracking characteristics of display devices describe the chromaticity variations depending on input pixel values for the achromatic and chromatic colours. A set of 8 red, green, and blue patches with an interval of 32 pixel values between them was

created for primary colours and achromatic colours. CIE 1976 u' , v' chromaticity coordinates of displayed patches were measured. The loci of each of the reproduced patches were plotted on u' , v' diagrams in Figures 3-9 and 3-10.

The chromaticities varied depending on the input pixel values. It was especially clear with pixel values below 64. This is a typical characteristic of LCD devices mainly due to inter-channel reflections and back light leak through the LCD filters (Fairchild and Wyble, 1998, Chou *et al.*, 2008). Non-zero black levels of 0.24 cd/m² and 0.15 cd/m² were measured on the CG210 and the CG245W, respectively. Further analysis of data was carried out by taking black levels into account using the model suggested by Fairchild and Wyble (Fairchild and Wyble, 1998). Colour tracking characteristics of both displays were plotted before and after the black level compensation in Figures 3-9 and 3-10, for the CG210 and the CG245W, respectively.

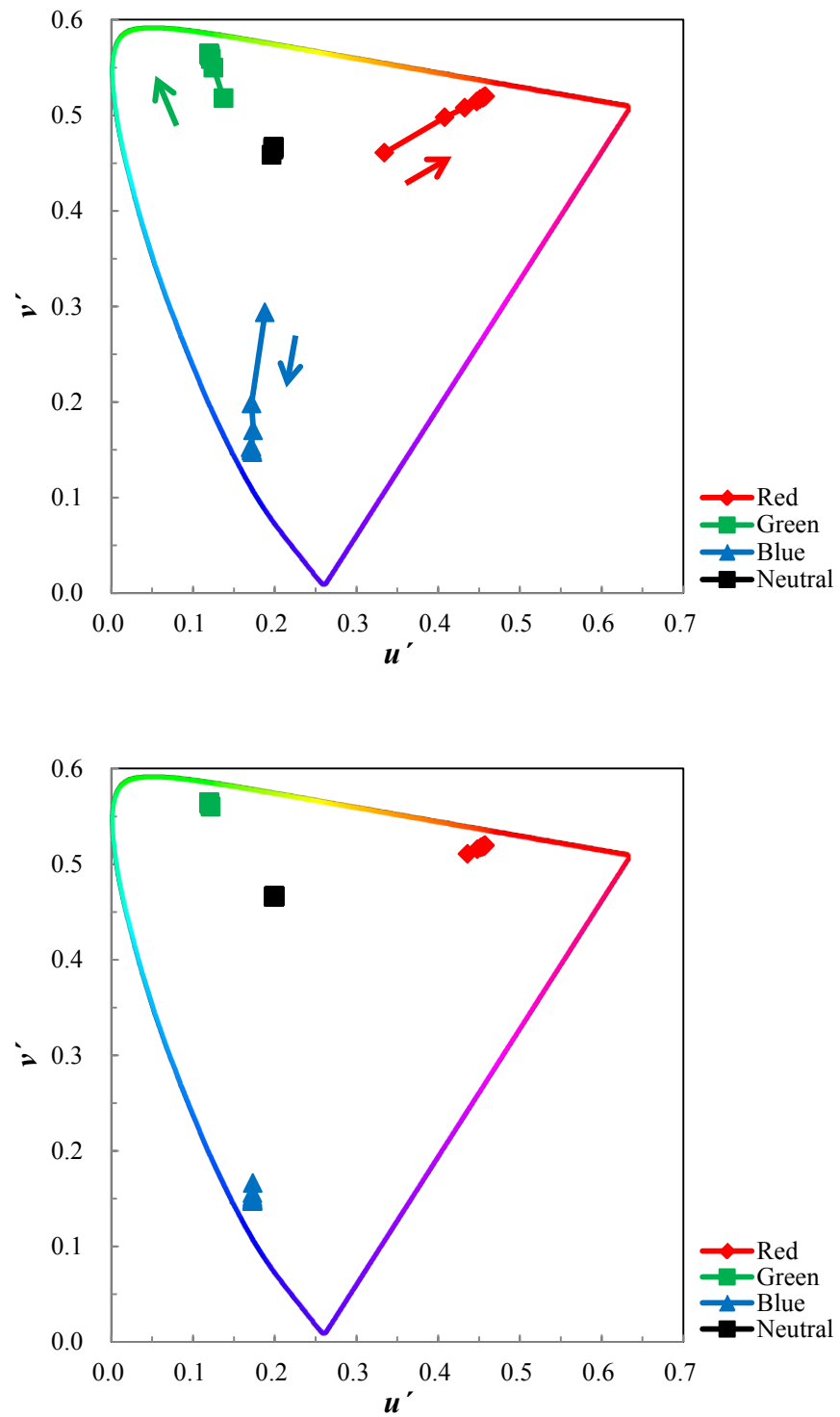


Figure 3-9. Colour tracking characteristics of the EIZO CG210, before (top) and after the black level compensation (bottom).

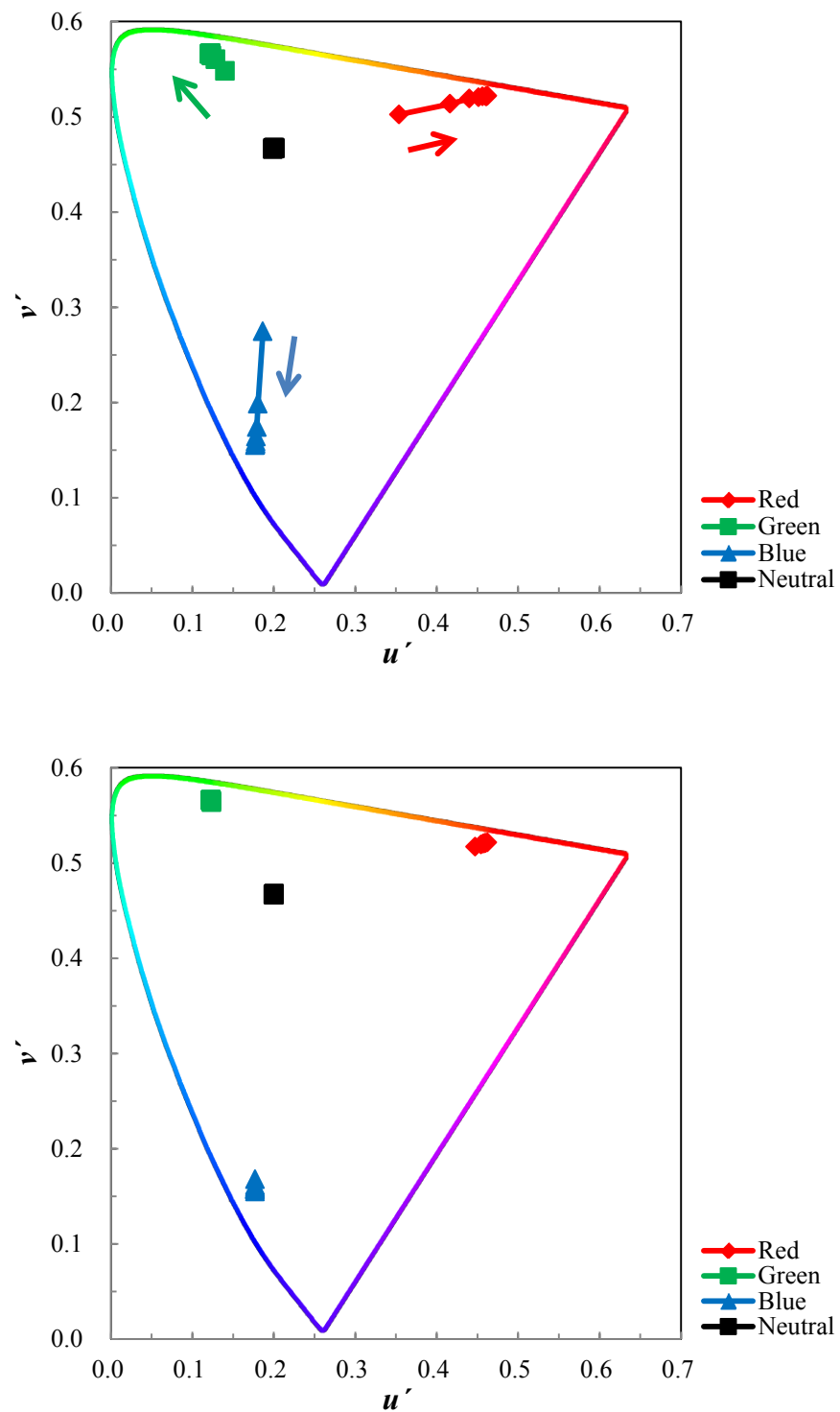


Figure 3-10. Colour tracking characteristics of the EIZO CG245W, before (top) and after the black level compensation (bottom).

3.2.5 Positional non-uniformity

Positional non-uniformity characteristics of display devices describe the variations in lightness and chromatic coordinates across the displayable area of an LCD screen. For the evaluation of the positional non-uniformity characteristics, the entire screen was filled with a white ($R=G=B=255$) patch. A total of 25 points were measured across the screen. The selected measuring points are illustrated in Figure 3-11.

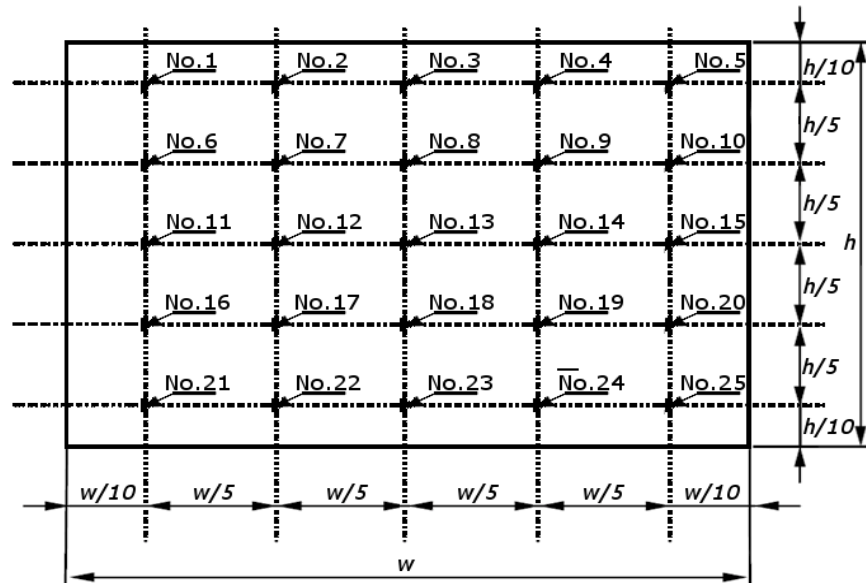


Figure 3-11. Positions of 25 selected points for positional non-uniformity characteristics of a display device. The ‘ h ’ and ‘ w ’ are height and width of the screen, respectively. Adapted from BS EN 61966-4 (*Multimedia Systems and Equipment--Colour measurement and management--Part 4: Equipment using liquid crystal display panels*. 2000, p.25).

CIELAB L^* , a^* , b^* values were measured and the differences between the reference point (No.13) and the measured points across the screen were calculated.

The variations in lightness, ΔL^* , across the screen were as big as 6.12 and 3.00 on the CG210 and the CG245W, respectively. On the CG210, the reference point (centre of the screen) was the brightest and the edges were measured to be darker. Lightness decreased as distance from the central region to the measured point increased.

The CG245 display, however, showed different characteristics. The top right area of screen was much brighter than the central region. The bottom area of the screen was the darkest. The results are shown in Figure 3-12.

The variation in chroma, ΔC_{ab}^* , across the screen was fairly large on the CG210 compared with that on the CG245W. The maximum chromatic variations, ΔC_{ab}^* , were 3.04 and 0.56 on the CG210 and the CG245W, respectively. The results are plotted in Figure 3-13.

In addition to the chromatic and lightness variations across the screen, the colour differences, ΔE_{ab}^* , were evaluated. Overall average colour difference, ΔE_{ab}^* , of 6.28 and 3.89 were found on the CG210 and the CG245W, respectively. Although the differences were not proportional to the distance from the reference point, position no. 13, the reproduction error, ΔE_{ab}^* , were generally higher at the edges of screen on both devices, as shown in Figure 3-14.

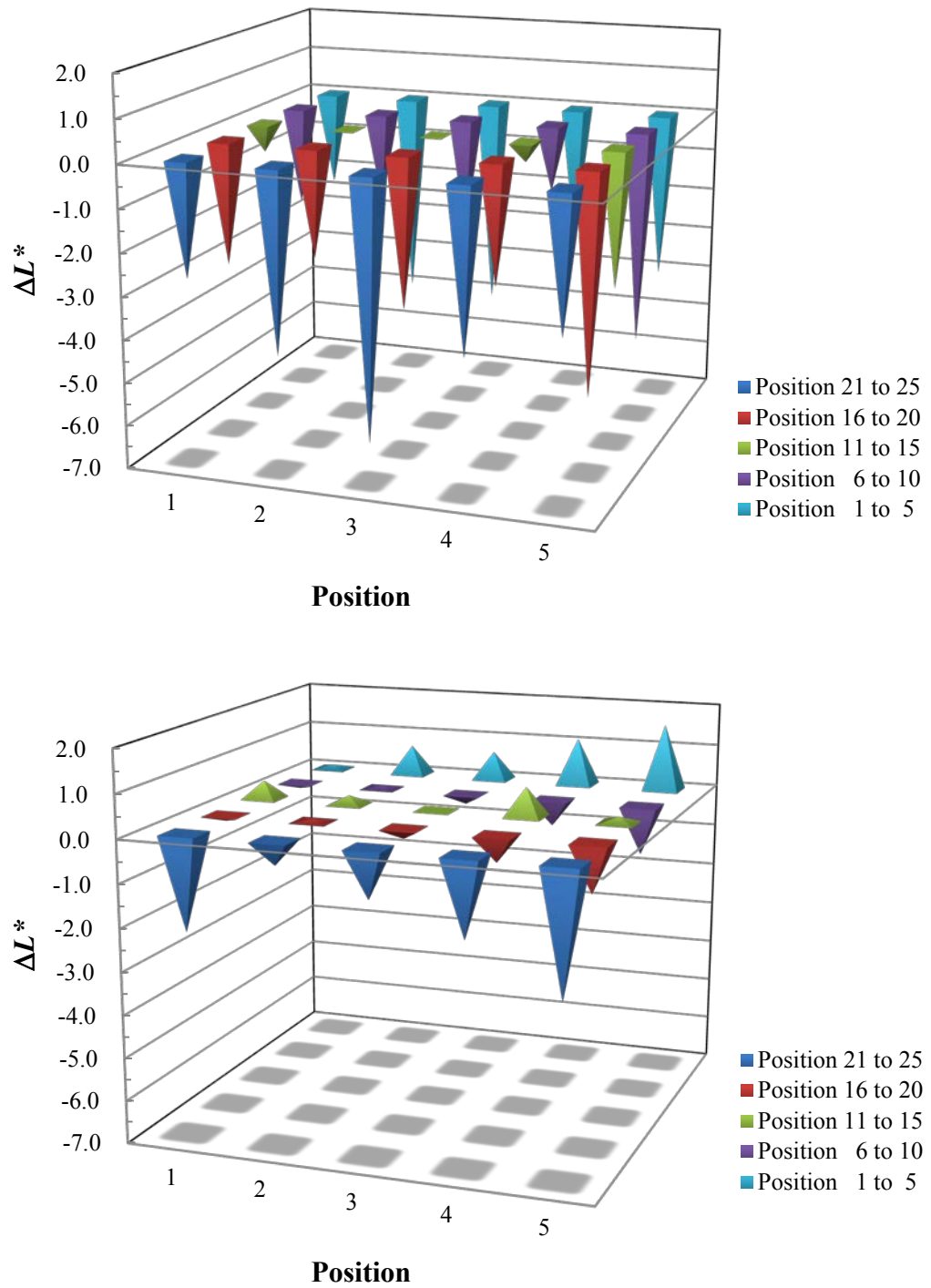


Figure 3-12. Lightness differences, ΔL^* , from the reference point to the measured points across the screen. The CG210 (top) and the CG245W (bottom).

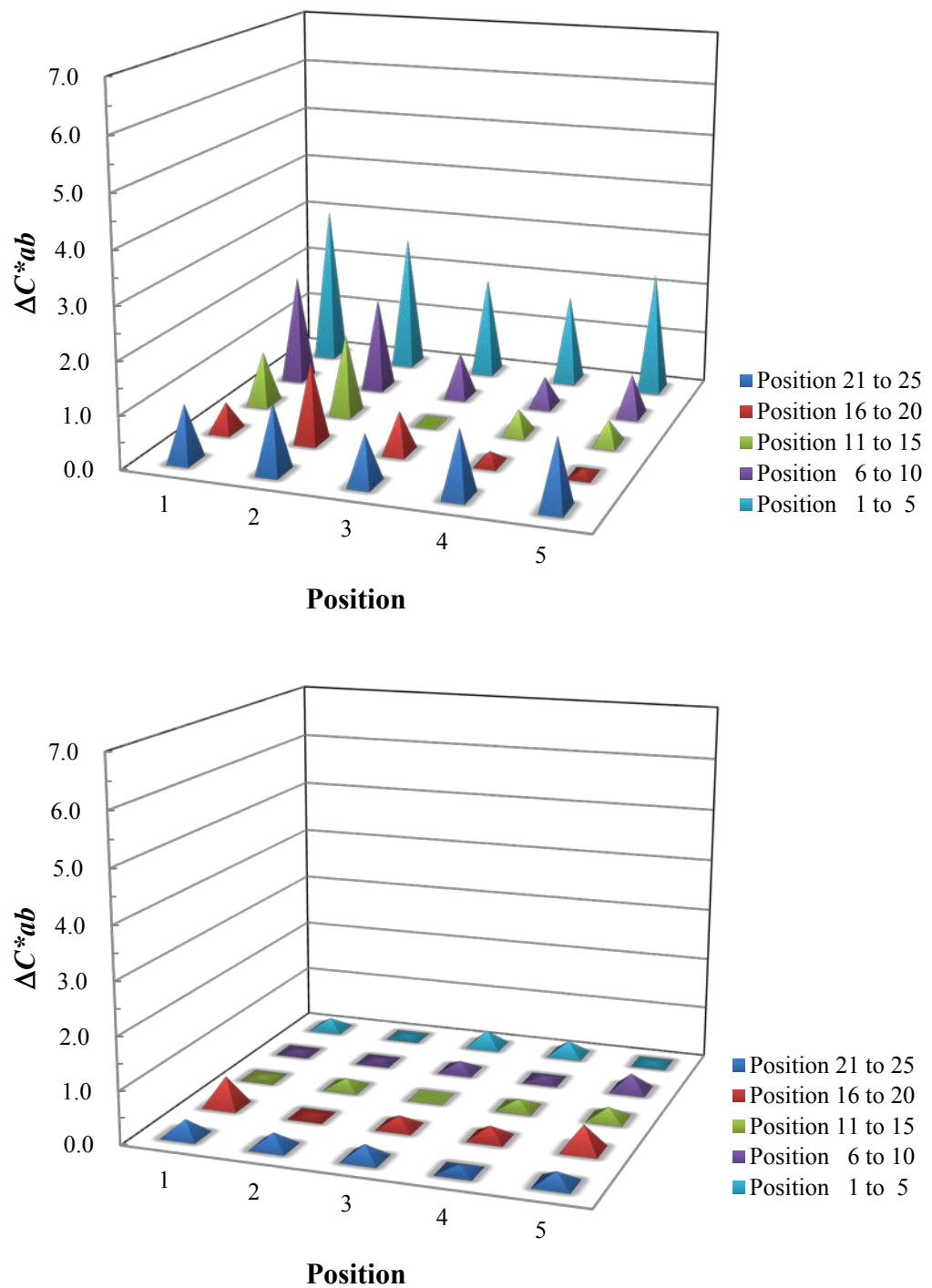


Figure 3-13. Chromatic differences, ΔC^*_{ab} , from the reference point to the measured points across the screen. The CG210 (top) and the CG245W (bottom).

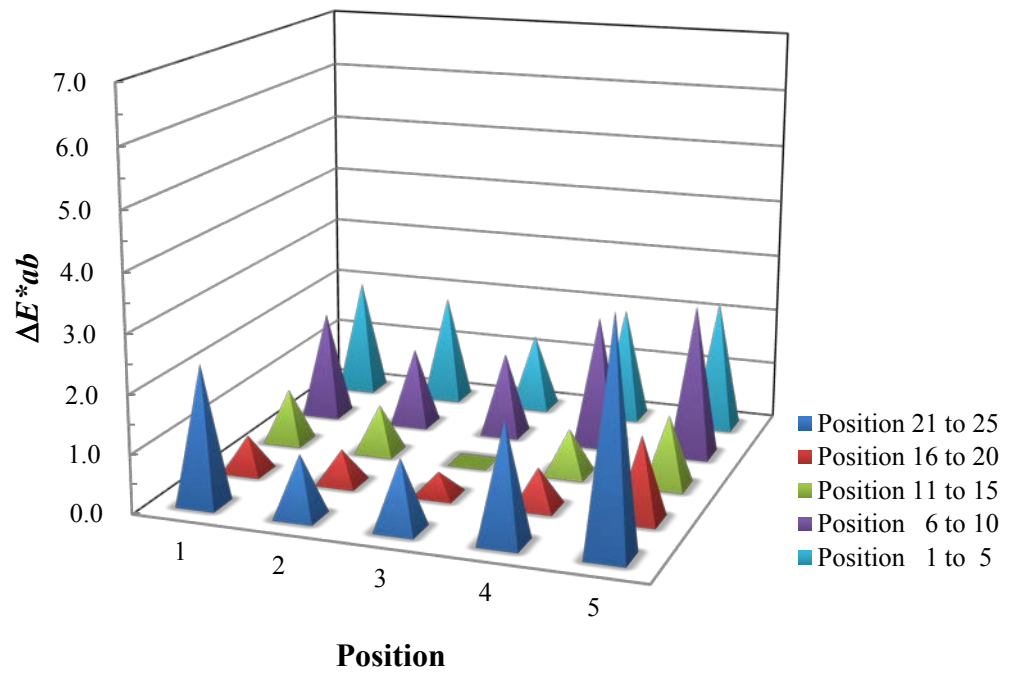
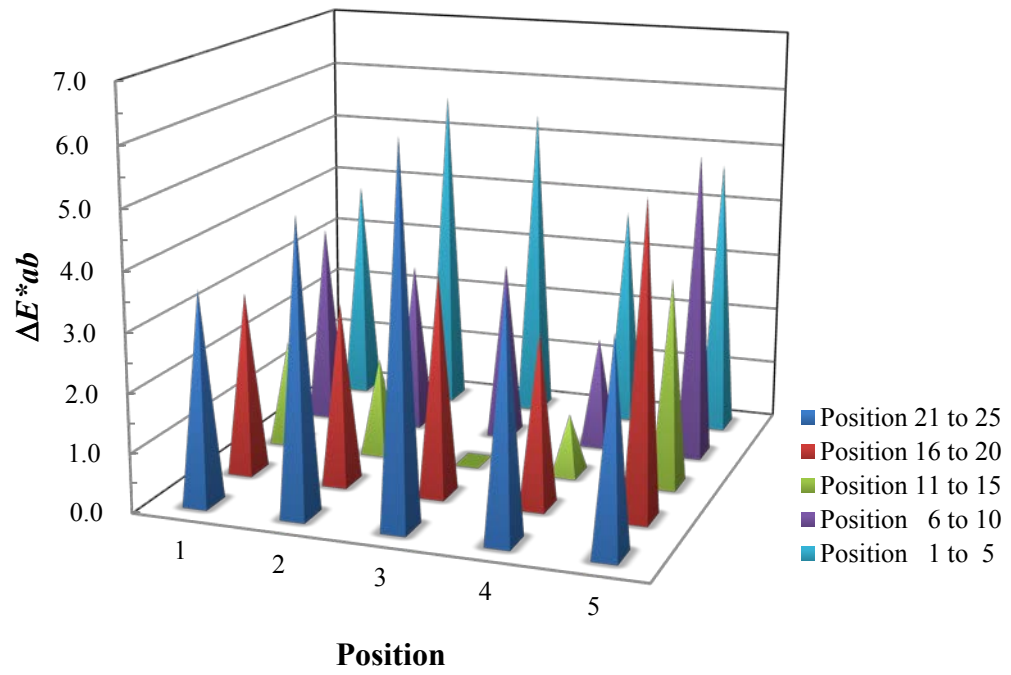


Figure 3-14. Colour differences, ΔE^*_{ab} , from the reference point to the measured points across the screen. The CG210 (top) and the CG245W (bottom).

3.2.6 Dependency on background

Dependency on background characteristics of display devices describes the effect of the background brightness on the centrally displayed patches or images. A pair of test patches was created. One contained a black background with a white patch in the central area, while the other contained a white on the entire patch.

CIELAB values of both patches were measured and the colour differences were calculated and shown in Table 3-5. From the results, it was found that the luminance and the chromaticity of the central measured region were independent from the background colour on both displays.

Display model	Black background			White background			ΔE_{ab}^*	ΔE_{00}^*
	L^*_1	a^*_1	b^*_1	L^*_2	a^*_2	b^*_2		
CG210	105.10	-7.33	-8.57	104.99	-7.55	-8.85	0.37	0.21
CG245W	106.79	-8.81	-8.12	106.73	-8.85	-8.26	0.16	0.18

Table 3-5. Measured CIELAB values and evaluated colour differences.

3.2.7 Temporal stability

Temporal stability characteristics of display devices describe the time required to reproduce stable output luminance and chromaticity of the display devices, and the variation in performance over period of time. Short-term stability characteristics for the duration of 2 hours with an interval of one minute and mid-term stability characteristics for the duration of 24 hours with an interval of ten minutes were investigated. Once the each display device was prepared to display a white patch, the display was turned off to cool down for minimum one day before the measurement. Then the displays were turned on and measurements were made one minute after the display was turned on.

Measurements were carried out for the duration of 2 hours for the short-term stability characteristics. For the mid-term stability characteristics, the first measurement was made 10 minutes after the display devices were turned on and the measurements were carried out for the duration of 24 hours.

The output luminance, Y (in cd/m^2), and the chromaticity coordinates x, y were measured and plotted against time in Figures 3-15 and 3-16. From the short-term stability characterisation, the CG245W display performed excellent with a standard deviation, σ , of 0.318 in luminance whilst the CG210 showed 5.279 after the first measurement. From the mid-term stability characterisation, the CG245W display performed excellent with a standard deviation of 0.191 in luminance whilst the CG210 showed 1.735. Both mid-term and short-term stability in chromaticities were, however, excellent with a standard deviation of less than 0.001 from both devices.

Both systems were calibrated and set to display at the peak luminance of $120\text{cd}/\text{m}^2$. However, the CG210 was found to be resetting the calibration settings automatically when powered off. It is also clear in Figure 3.15 that the output luminance of the EIZO CG210 fluctuated even after a long warm up time.

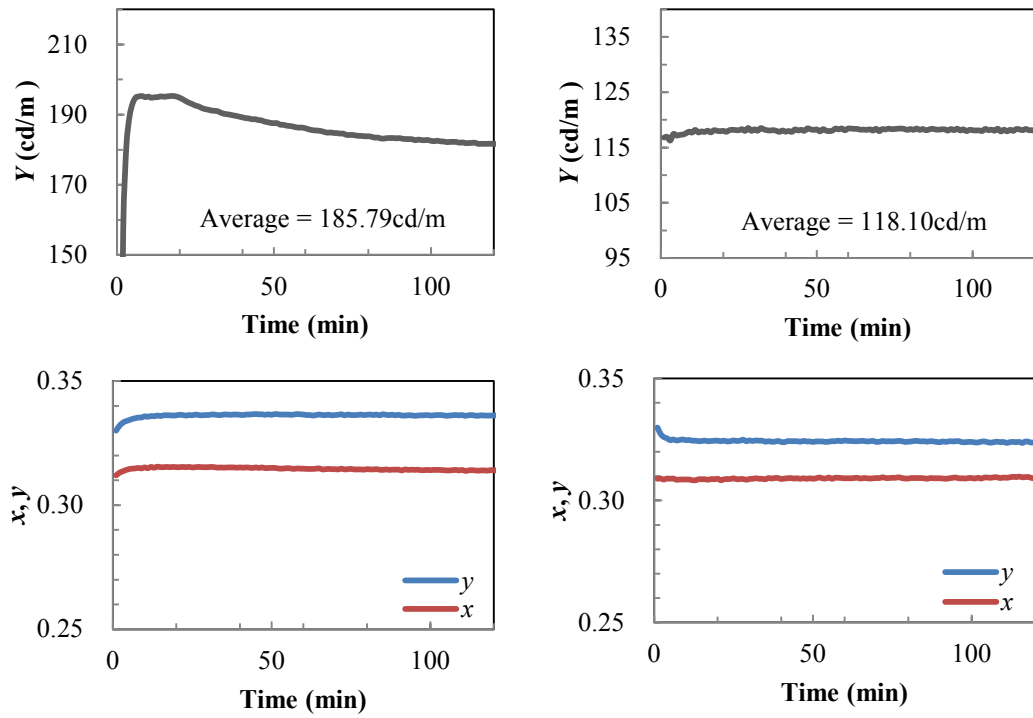


Figure 3-15. Short-term stability in luminance (top) and in chromaticities (bottom), on the CG210 (left) and on the CG245W (right).

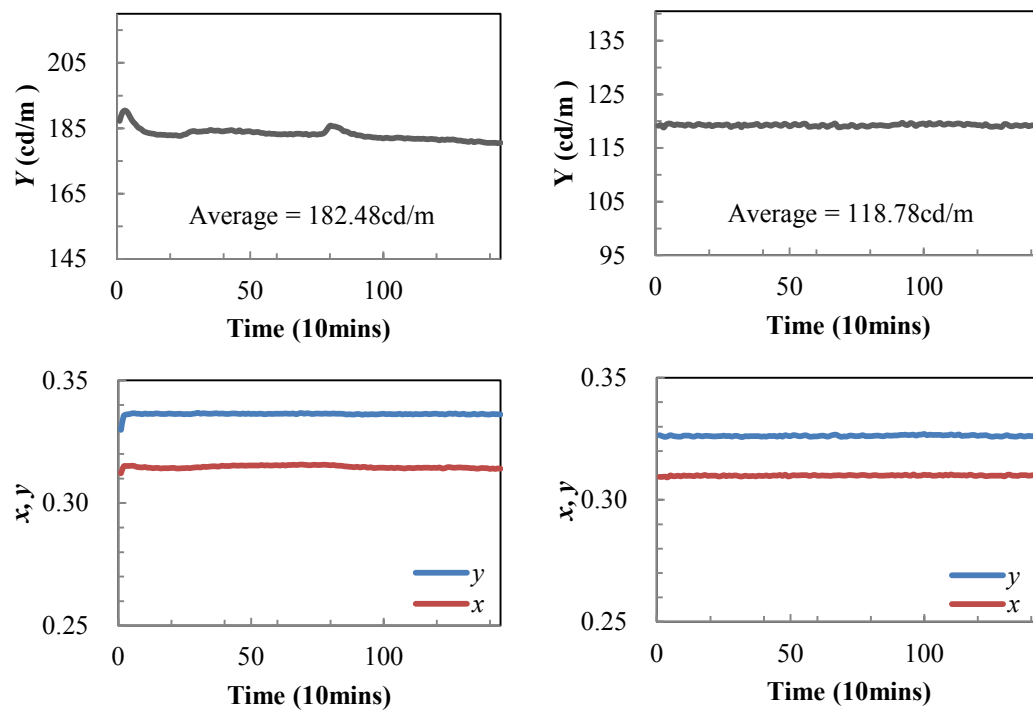


Figure 3-16. Mid-term stability in luminance (top) and in chromaticities (bottom), on the CG210 (left) and on the CG245W (right).

3.2.8 Viewing angle dependency

Viewing angle dependency characteristics of display devices describe the effect of the viewing angle on the output luminance and chromaticity. Both devices were equipped with a tilt stand which allowed changes in the vertical viewing angle. A turntable device which allowed accurate horizontal swivel was placed under the display stand for the evaluation of horizontal viewing angle characteristics.

A set of test patches containing 8 neutral and 3 pure primaries in the centre area, with a black background, were prepared. Angular dependency on luminance, Y , as well as CIE 1976 u' , v' chromaticity coordinates over a horizontal range of centre $\pm 40^\circ$ at an interval of 10° were measured. The $(-)$ angles represent viewing from the left side and the $(+)$ angles represent viewing from the right. The measurement was repeated for a vertical range of -5° and $+20^\circ$ for the CG245W and a vertical range of 0° and $+20^\circ$. The angular measurement distance interval was 5° .

The changes in output luminance at various viewing angles are plotted in Figure 3-17. The variations in chromaticities are also plotted on chromaticity diagrams in Figure 3-18. It was clear that the loss in luminance was fairly large on both devices. The loss in luminance was slightly higher when the viewing angle was changed vertically. However, the changes in chromaticities at different viewing angle were fairly small on both devices.

In addition to the pure primary colours and the white, a set of neutral patches were measured and plotted in Figure 3-19.

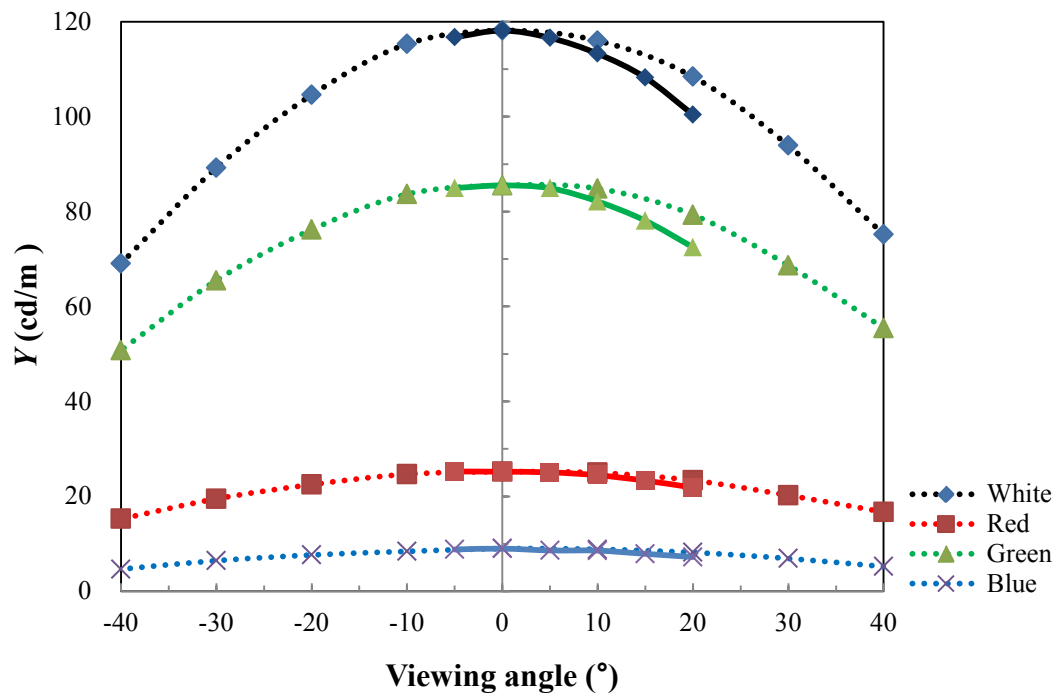
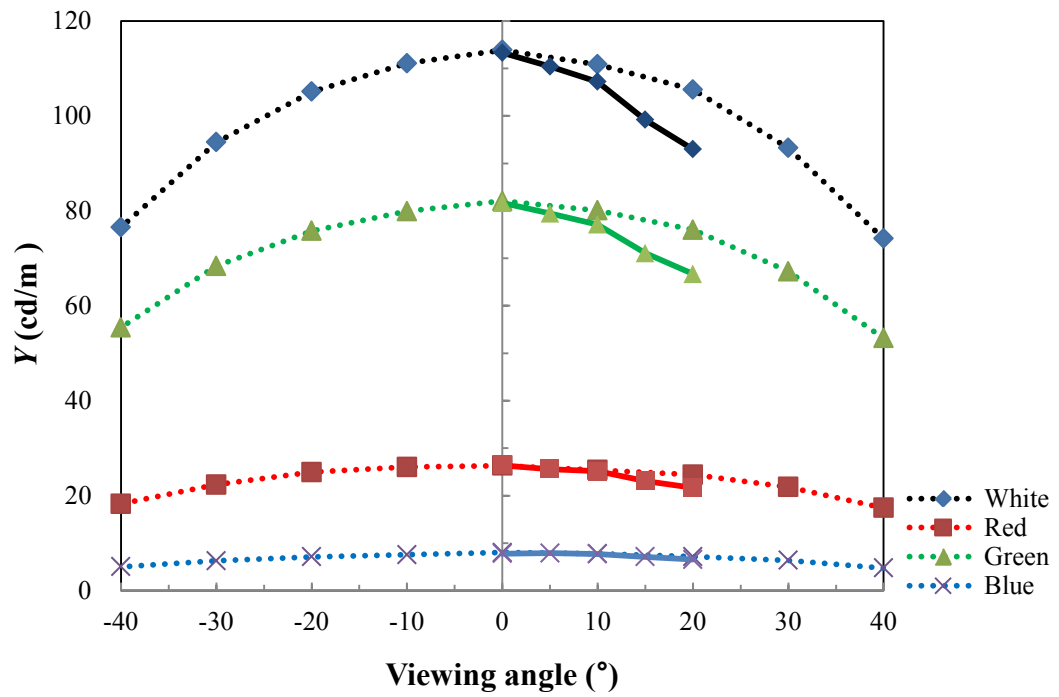


Figure 3-17. Luminance output of the pure primaries and the white at various horizontal and vertical viewing angles. Solid lines represent vertical luminance and broken lines represent horizontal luminance. The CG210 (top) and the CG245W (bottom).

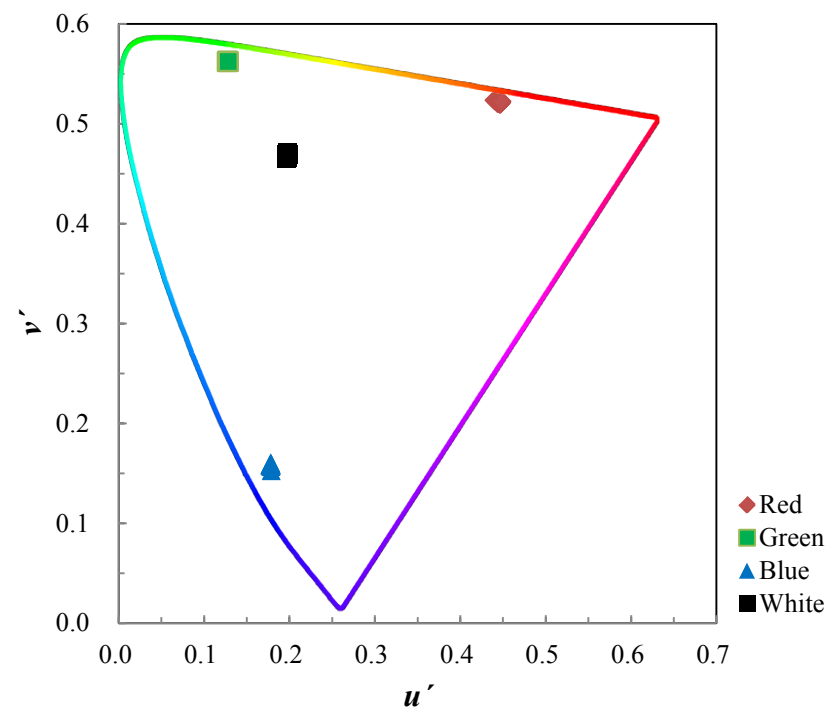
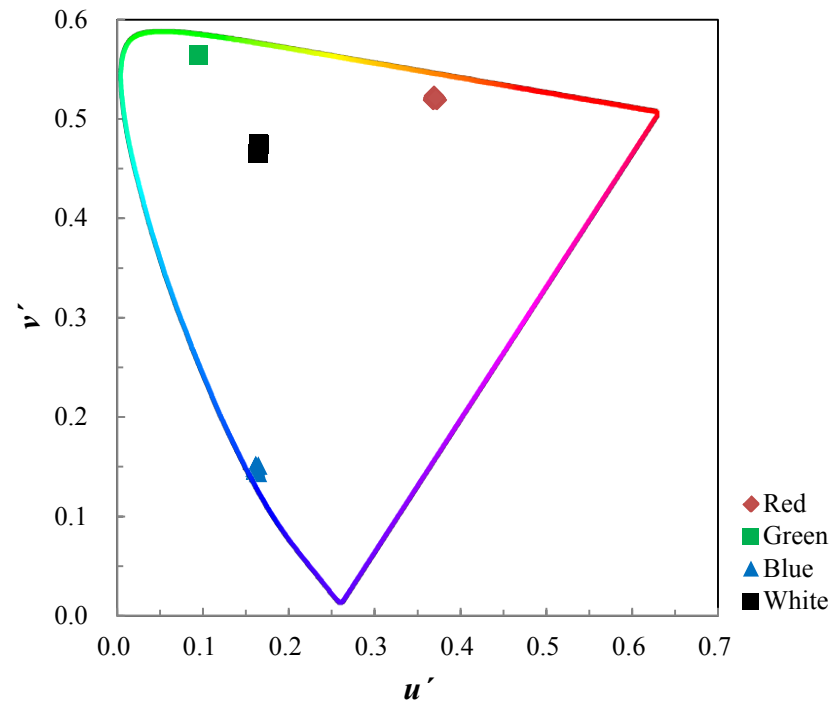


Figure 3-18. Changes in chromaticities at various viewing angles. The CG210 (top) and the CG245W (bottom).

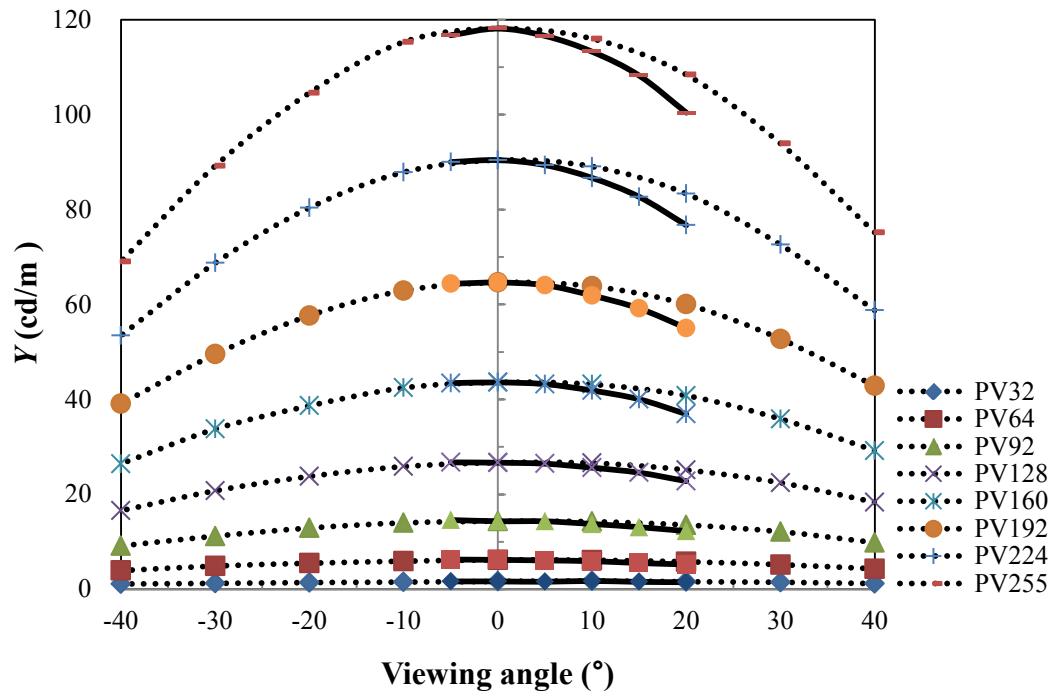
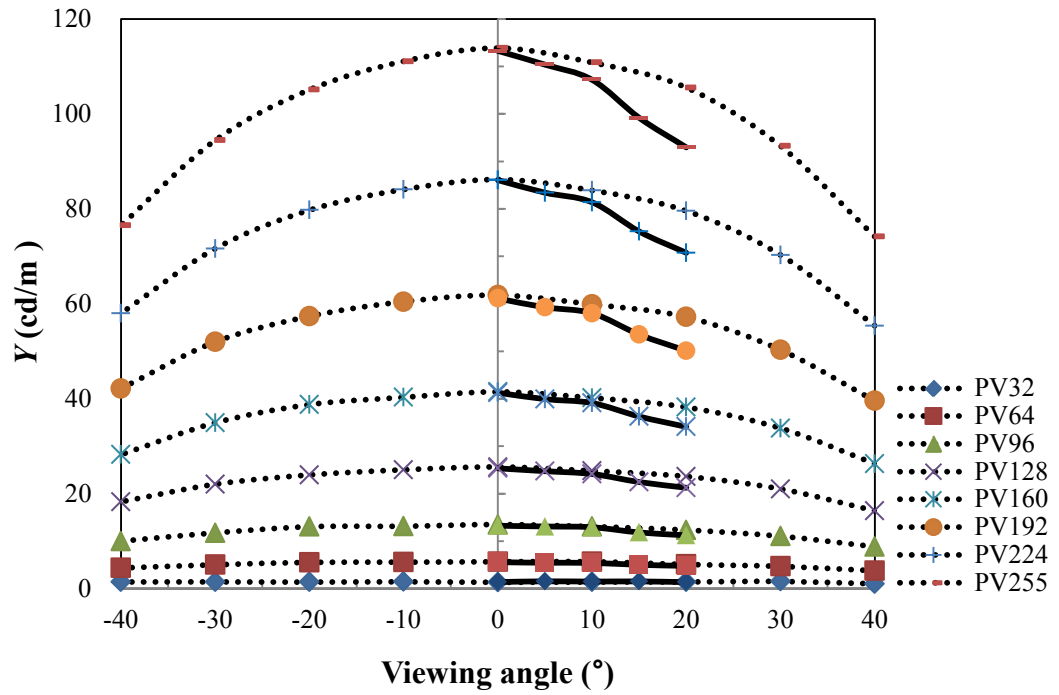


Figure 3-19. Changes in luminance output of neutral patches at various horizontal and vertical viewing angles. Solid lines represent vertical luminance and broken lines represent horizontal luminance. On the CG210 (top) and on the CG245W (bottom).

3.2.9 Positional non-uniformity at the observation plane

All of the above characteristics were measured on the plane of the display faceplate. However, the psychophysical investigations were carried out by observations made at a certain viewing distance and in a plane of observation parallel to the display. The size of a standard reference image had a horizontal visual angle of approximately 20 degrees at the set observation distance. In order to investigate the impact of the display characteristics on the psychophysical experiments in the following chapters, the positional non-uniformity of the CG245W display was investigated at the observation plane.

A Konica-Minolta CS-200 tele-chroma meter was placed 60cm away from the centre of the display device, in a plane parallel to the display to mimic the position of the observations. The entire screen was filled with white ($R=G=B=255$). A total of 13 points (position 6 to 20, covering the display area used in Chapters 5 and 6) were measured from the observation plane by tilting and swivelling the measuring instrument (c.f. Section 3.2.5).

Results showed that every position measured was darker compared with the central reference position (No.13), the differences in lightness, ΔL^* , ranging from 0.20 to 4.11. The differences in lightness were larger at the horizontal orientation than at the vertical orientation. Also, the differences in chroma, ΔC_{ab}^* , were evaluated to be higher (maximum of 1.45) than the results obtained from positional non-uniformity characterisation (maximum of 0.56). Overall, the colour differences, ΔE_{ab}^* , were calculated and plotted in Figure 3-20. The average error was 2.36 with a ranged from 0.97 to 5.24.

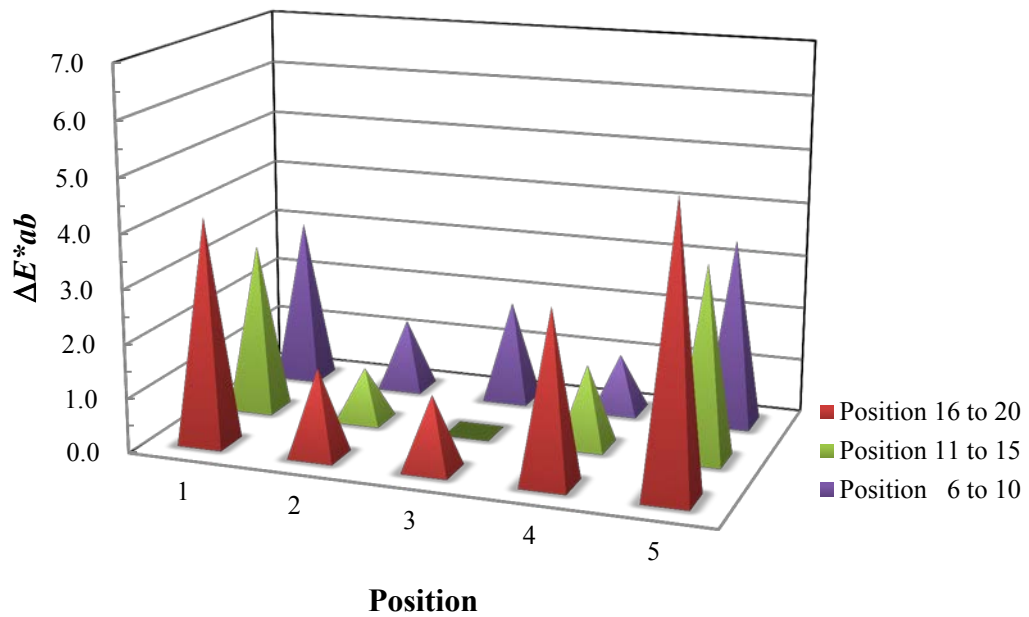


Figure 3-20. Colour differences, ΔE_{ab}^* , from the reference point to the measured positions across the screen.

3.2.10 Summary

The display devices were characterised by adapting the methods suggested in BS EN 61966-4. Two liquid crystal displays from the same manufacturer were used in this project, purchased at different time. The EIZO CG210 was originally used in the investigation described in Chapter 4. The CG210 exhibited a good black level, less than 0.2% of maximum luminance, compared with the typical LCDs 0.4% (Fairchild and Wyble, 1998). Also, the primary colours at full strength were reproduced accurately at the central region of the screen. However, the colour reproductions across the screen were non-uniform. The maximum differences in chroma, ΔC_{ab}^* , and in lightness, ΔL^* , were as big as 3.04 and 6.12, respectively, which resulted the colour reproduction errors, ΔE_{ab}^* , ranging from 1.02 to 6.28. The differences were especially bigger at the

edges of the screen; however, it was not proportional to the distance from the reference point (centre of the screen). After the characterisation of the display, it was found that the positional non-uniformity of the display not sufficiently uniform for further investigations (c.f. Chapters 5 and 6). Therefore, the EIZO CG245W was purchased at the stage of the project for use in the matching experiments.

The CG245W exhibited black levels of less than 0.1% and accurate reproduction of tone and colours at the central region of the screens. Also, the CG245 display showed a good temporal stability. The standard deviations of the output luminance, were $\sigma=0.27\%$ over a period of 2 hours and $\sigma=0.16\%$ over a period of 24 hours. However, the CG245W exhibited slight variations, just over a perceptible limit (Berns *et al.*, 1993), of its performance across the screen. Average reproduction error, ΔE_{ab}^* , was 1.53, ranging from 0.38 to 3.89. This was mainly due to the variations in lightness.

Further, the CG245W also showed considerable angular dependency on luminance. However, the observations were made at set distance without changing the angle of display plane, the viewing angle dependency characteristic evaluated by tilting and swivelling did not have significant meaning. Therefore, the positional non-uniformity characteristic of the limited display area, where the test images are displayed during the visual investigations in Chapters 5 and 6, was evaluated from the position of the observation to mimic the position of the observations. The colour differences, ΔE_{ab}^* , were found to be higher with an average of 2.53, ranging from 0.97 (near central reference point) and 5.24 (at the edges); it was relatively proportional to the distance. Even though the colour reproduction of the CG245W was not spatially independent, the errors were below the commonly accepted limit (Abrardo *et al.*, 1996).

Since the display device was used to display a pair of images side by side for the visual matching experiments in the further experimental work in Chapters 5 and 6, the positional uniformity characteristic of the display was main concern, limiting the display systems' performance. By repeating the pair of displayed images in random order each time, the experiments were carried out, the impact of the slight display non-uniformity was minimised (Jin *et al.*, 2009). The various characteristics of the display devices evaluated in this chapter were considered for the determination of the final choice of devices, the software preparation and interface design (i.e. determination of the display areas, randomisation of the displayed image positions, etc.).

Chapter 4

Psychophysical investigation 1:

Identification of image attributes that are most affected by changes in displayed image size

This chapter is concerned with the investigation of changes in image appearance when images are viewed at different image sizes on an LCD device. Aim of this psychophysical investigation was to identify image attributes that were most affected visually by changes in displayed image size. This was achieved by collecting data from a series of visual experiments using the rank order method to obtain ordinal scales. This chapter first describes the preparation of test stimuli by the image capture, the selection, and the image processing. Secondly, it briefly describes the experimental set up and the

method employed. Finally, the results were further discussed in relation to the scene characteristics. Further, research was carried out to link original scene content to the attributes that changed most with changes in image size.

4.1 Preparation of test stimuli

4.1.1 Image capture

Two digital image capturing devices (with built-in LCDs) of different overall image quality were used for recording identical natural scenes with a variety of pictorial contents. An eight megapixel Canon 30D digital SLR camera, equipped with an EF-S10-22mm lens and a two megapixel Apple iPhone camera with a built-in lens were used for capturing the test images. Both cameras were characterised by the methods described in Chapter 3. Camera specifications and settings employed during scene capture are also explained in Chapter 3.

For each captured scene both devices were set to the same focal length to record ‘identical’ image frames. Because shutter speed and lens aperture are not adjustable manually on the Apple iPhone camera, several exposures were made for each scene with the Canon 30D camera in manual mode to visually match the captured image by the Apple iPhone. As the Apple iPhone allowed saving captured scene as JPEG files in sRGB 8-bit per channel colour coding only, all images captured by both cameras were saved in JPEG format (*Information technology--Digital compression and coding of continuous-tone still images: Requirements and guidelines*. 1994). The most similarly exposed images from both cameras were selected by visual inspection to create the appropriate test set.

4.1.2 Image selection

Image selection was carefully carried out, to include scenes representative of various possible situations and conditions from ordinary digital camera users. For each capturing device, a total of sixty-four captured scenes, including architecture, nature, portraits, still and moving objects, and artwork under various illumination conditions and recorded noise levels were selected. The selected test sets included some colourful and some rather neutral images, images that contained sharp edges, or unsharp (out of focus) edges resulting shallow depth of fields, images with more or less fine detail, some relatively dark and some light images. The purpose for this variation on test image content was to investigate the relationship between groups of images with varying characteristics and their appearance changes with changes in the displayed image size. For the investigation of the effect of motion blur on image appearance, the test sets included some images where camera shake was purposefully introduced.

4.1.3 Image processing

The original captured images from both capturing devices were too large to be displayed at full resolution on an EIZO ColorEdge CG210 21.3" LCD, which was used in this investigation. Thus, the test images were sub-sampled from their original sizes to 744(H)×560(V) pixels, using bi-cubic interpolation. The effect of the interpolation on the Spatial Frequency Response (SFR) of the images was investigated; it is presented in Chapter 5. The size of the sub-sampled reference images was approximately half of the LCD's native horizontal and vertical pixel resolutions. It is an appropriate image size to be displayed on commonly used displays. Achromatic versions of the test stimuli were also prepared to investigate various image attributes, such as contrast, sharpness,

brightness and noise, which are affected mostly by the luminance (achromatic) channel (Hunt, 2004, p.48-59). The achromatic versions of the test stimuli were obtained using Adobe Photoshop v.7.0, by converting all sub-sampled images from sRGB to CIELAB space and selecting the lightness channel (L^*) for the purpose.

4.2 Psychophysical investigation

Due to difficulties in controlling the ambient lighting in the laboratory, the experimental work was conducted in a totally dark environment. Although the reference sRGB display viewing conditions are dim (ambient illuminance of 64lux, and veiling glare of 0.2 cd/m^2) (*Multimedia Systems and Equipment--Colour measurement and management--Part 2.1: Default RGB colour space--sRGB*. 2000), the advantage of conducting experiments in such an environment was that the display was free from veiling glare, which is known to decrease the perceived contrast and colour saturation (Hunt, 1952). Displaying images in dark rather than in dim conditions produced a slightly reduced overall image contrast (Chapter 4 of Fairchild, 2005). This variation was not considerable and did not affect the visual quality of the original test images.

The rank order method was chosen and implemented in this chapter. It is the quickest and most straight forward for obtaining ordinal data (Engeldrum, 2000, p.79). As we have discussed in the earlier section (c.f. 2.3.3.2), the main disadvantage of the ordinal scale is that it does not possess any significance of differences. However, the aim of this investigation was to find which attribute(s) were most affected by displayed image size. The magnitude of differences was secondary importance, since we planned to further evaluate the perceptual differences in later research (c.f. Chapters 5 and 6).

4.2.1 System calibration and settings

The EIZO ColorEdge CG210 21.3” LCD, presented in Chapter 3, driven by a Sony VAIO VGN-T92S computer with an on-board graphics controller, was used in the psychophysical investigation (c.f. Section 3.2.10). Calibration was achieved using the GretagMacbeth Eye-One Pro with Profilemaker v5.0 to the settings presented in Chapter 3. Display calibration was repeated daily throughout the period of the psychophysical investigations.

4.2.2 Software preparation and interface design

The psychophysical display application was designed to display an image in two different displayed image sizes, side by side. It was written in JavaScript and optimised in Mozilla Firefox v3.0.1 web browser (Mozilla, 2013). A mid-grey (50% luminance) background was selected; at a display gamma of 2.2, it corresponded to a pixel value of 186 for all three R, G and B channels. This mid-grey was selected to minimise background effect on the appearance of test images (Choi *et al.*, 2007b, Choi *et al.*, 2007a).

During the experiment, each test image was displayed simultaneously at two different sizes; one at the “original” control size of 744(H)×560(V) and the other equivalent to the size of the built-in LCDs of each capturing device (186(H)×140(V) for the Canon 30D and 244(H)×182(V) for the Apple iPhone). Test images were displayed in random order and in random left-right display positions, (i.e. left: large image, right: small image, or vice versa). The application automatically recorded the observation data and saved them as a text file on the computer’s hard disk. The display interface is illustrated in Figure 4-1.

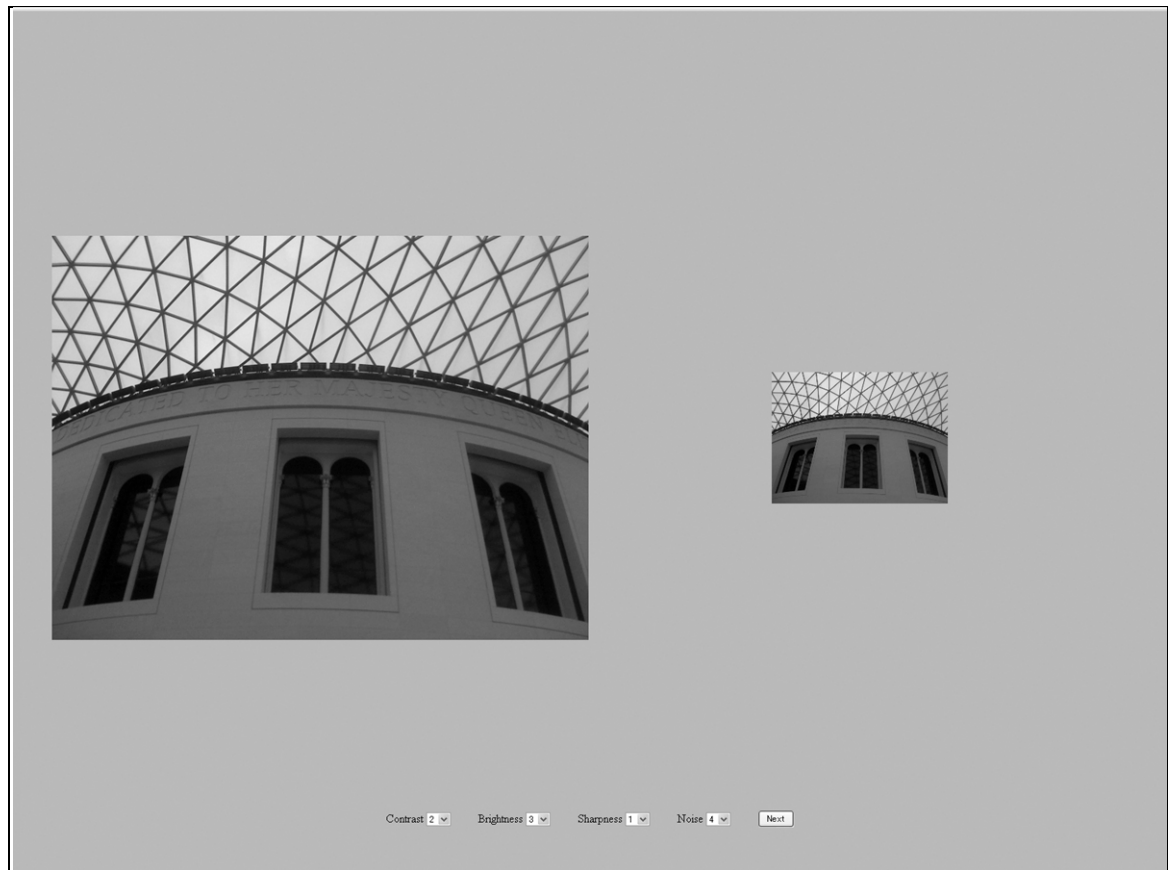


Figure 4-1. Display interface for the psychophysical test page in achromatic mode.

4.2.3 Rank order method

Rank order experiments were conducted by displaying the same image at two different sizes, side by side. The display interface is illustrated in Figure 4-1. Each observer took the test four times: for each camera, they judged both the achromatic and the chromatic versions of the test stimuli. Observers were seated approximately 60cm away from the display to keep the angle of subtense and were asked to sustain the viewing distance. However, observers were not forced to keep it with a chin-rest, which was used in later experiments (c.f. Chapters 5 and 6). At the set distance, visual angle of the reference image was approximately 20 degrees. Observer's instructions were provided to the observers to rank-order the attributes that were affected with changes in displayed image size. For the achromatic version of the stimuli they rank ordered the following

attributes (from 4 being the most affected to 1 being the least affected): contrast, brightness, sharpness and noisiness. For the chromatic versions, in addition to the previous attributes, they also rank ordered hue and colourfulness (rank order from 6 to 1, including the previous four attributes). A total of seventeen observers, 9 females and 8 males, took the experiment. Their age ranged between 20 and 60 years old and had normal, or corrected, visual acuity. They had all previously been tested for colour deficiencies and they were mostly from imaging and design backgrounds.

4.3 Classification of test images

The original full size versions of the test stimuli were categorised by both an objective method, suggested by Triantaphillidou *et al.* (Triantaphillidou *et al.*, 2007), and by visual inspection. The authors have suggested a method to analyse and classify scenes by deriving certain *scene metric values* that identify how much or how little of a ‘scene characteristic’ relevant to an image quality attribute the image may possess. In brief, to discuss the proximity of the scene metric values between the different scenes and therefore the similarity in their characteristics, they classified these scene metrics into four ranges, ordered by relative distance from the median value quantifying the ‘scene characteristic’. Values found to be within one standard deviation, σ , from the median value were considered ‘average’ and were split into ‘average-to-low’, if below the median and ‘average-to-high’, if above the median. Another two categories included ‘extreme’ values, i.e. values that were more than $\pm 1\sigma$ away from the median, comprising the ‘very low’ category if more than 1σ below the median, and ‘very high’ if more than 1σ above the median.

To investigate the relationships between the attribute rankings and the test-scene content, the test stimuli in this work were classified according to five different image characteristics: three characteristics were affected by both scene content and the capturing system's performance and two further characteristics were affected solely by the system's performance. For examining the lightness (i.e. how light/dark they may look to the observer) and the colourfulness (i.e. how colourful/non-colourful they may look) of the test images, the original stimuli were classified using objective scene analysis. In addition, all pairs of stimuli (same imagery with both cameras) were visually inspected to investigate the effect of system performance on the classification. For examining the busyness, sharpness and noisiness of the stimuli, only visual inspection in a dark surround was used.

Table 4-1 shows the selected image characteristics and the number of images that fall in each category, for both the Canon 30D and the Apple iPhone cameras. The median CIELAB L^* values were used to classify image lightness. The test images obtained by both devices were similarly classified, i.e. similar number of images in each category. The variance in chroma of the images, VC_{ab}^* , which has been shown to correlate with the perceived colourfulness (Triantaphillidou *et al.*, 2007), was used for the 'colourful/non-colourful' classification. In this case, images were classified differently for each device, due to differences in the capturing system performance. For example, some images from the Apple iPhone camera were objectively classified as 'colourful', simply due to high colour noise levels, even if they were not otherwise especially colourful in visual appearance.

The test images were finally categorised for busyness, sharpness and noisiness, by visual inspection. Careful subjective classification was carried out using a calibrated

to sRGB monitor and similar viewing conditions to these employed in the psychophysical tests. For the busyness characteristics, images possessing high amount of detail and texture were categorised as ‘busy’, whereas those possessing mainly slowly varying areas were categorised as ‘non busy’. For the sharpness characteristics, images that were in sharp focus and/or possessing sharp edges were categorised as ‘sharp’, whilst those that appeared out-of-focus and/or possessing blurred and moving objects and/or where camera shake had been introduced during capture, were categorised as ‘un-sharp’. For the noisiness characteristics, the test images appeared to possess any visible noise were categorised as ‘noisy’. The number of images assigned in each category for latter two characteristics were different for the two camera systems, since sharpness and noisiness are highly dependent on imaging system performance (Triantaphillidou, 2011a, p.348).

	Dark/Light				Colourful				Busy		Sharp		Noisy	
	Dark	Moderately dark	Moderately light	Light	Colourful	Moderately colourful	Moderately non colourful	Non colourful	Yes	No	Yes	No	Yes	No
Canon 30D	8	24	22	10	17	15	21	11	30	34	49	15	30	34
Apple iPhone	8	24	21	11	21	11	11	21			43	21	62	2

Table 4-1. Images classified according to their lightness, colourfulness, busyness, sharpness and noisiness.

4.4 Results and discussion

The ranked data for all test stimuli were averaged for each mode (achromatic and chromatic) and each camera. The attribute with highest average rank means ‘the most affected attribute’ with changes in the display image size. Figure 4-2 shows the average ranks with standard error bars for all images captured by both cameras. The attributes

with the higher average rank mean that the corresponding attributes are affected more by changes in the displayed image size.

Overall, for the achromatic version of the stimuli obtained by both cameras, observers ranked sharpness as the most affected attribute and contrast as the second most affected attribute, by changes in the displayed image size. Brightness and noisiness were ranked third and fourth, respectively. However for the stimuli obtained using the Apple iPhone camera, the difference in average ranks for brightness and noisiness became very small, with 2.19 for brightness and 2.20 for noisiness. In Figure 4-2, the variations in the average ranks were similar for contrast and brightness with approximately 0.3 for both stimuli sets. However, it was clear that the variations in the average ranks for sharpness and noise were higher for stimuli obtained by Apple iPhone camera compared with those obtained by the Canon 30D camera.

For the chromatic version of the stimuli captured with the Canon 30D, observers again ranked sharpness, contrast, and brightness in the same order (i.e. first, second and third) as for the achromatic stimuli. Colourfulness and hue were found to be affected less, whilst noisiness was the least affected attribute, although it is to be noted that most test images did not contain significant amounts of noise.

For the chromatic version of stimuli from the Apple iPhone camera, observers again ranked sharpness and contrast as the two most affected attributes. However, noisiness, colourfulness and brightness were ranked as third, fourth and fifth whilst hue was ranked last.

The ranking of the attributes for the chromatic version of stimuli from the Apple iPhone camera was different, with noise and colourfulness having higher ranks, compared to the stimuli from the Canon 30D. This result is related to the relatively

lower average lightness and the higher level of chromatic noise that were present in most of the test stimuli originating from the Apple iPhone camera.

Overall, hue was affected less, or not at all with changes in displayed image size. This was also found by Xiao *et al.* in various colour appearance experiments (Xiao *et al.*, 2003, Xiao *et al.*, 2004).

Overall rank-orders of the attributes are as below:

Achromatic stimuli: Sharpness > Contrast > Brightness > Noisiness (Both devices)

Chromatic stimuli: Sharpness > Contrast > Brightness > Colourfulness > Hue > Noisiness (Canon 30D)

Sharpness > Contrast > Noisiness > Colourfulness > Brightness > Hue (Apple iPhone)

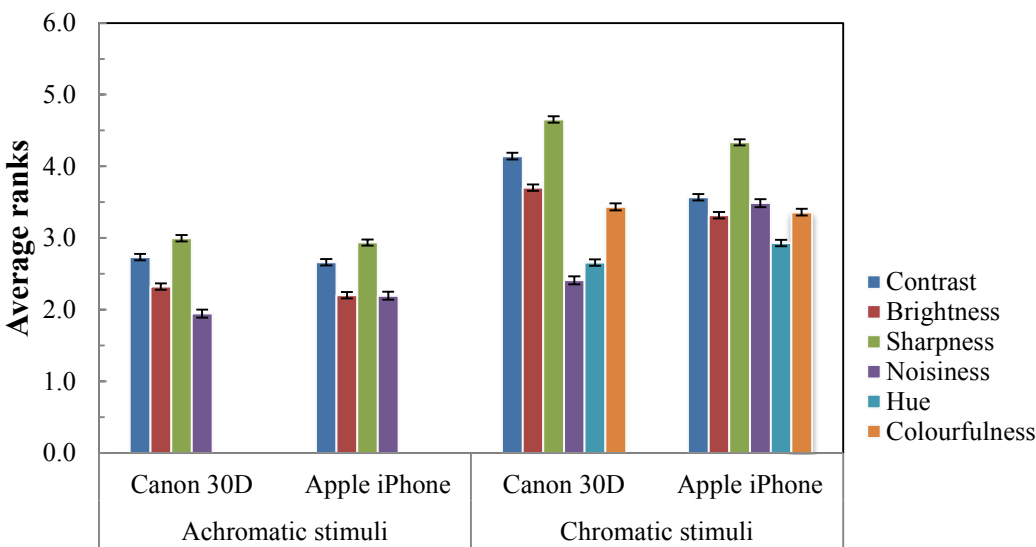


Figure 4-2. Average ranks from all test stimuli.

Figures 4-3 to 4-7 present the average ranks in relation to image categories that were described in Section 4.3.

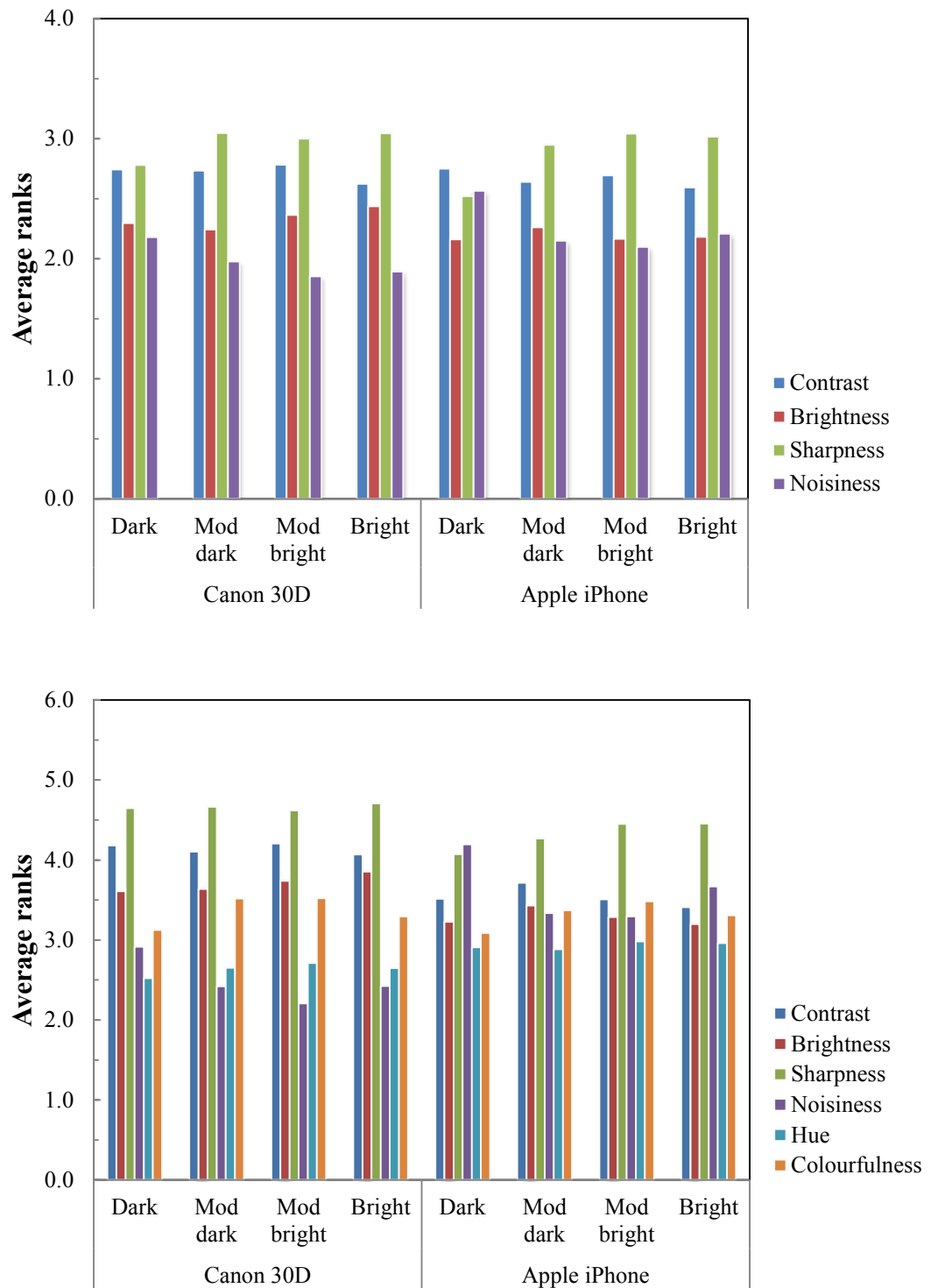


Figure 4-3. Average ranks of the image attributes of test stimuli categorised by their average lightness. Results from the achromatic versions of stimuli (top) and from the chromatic versions of stimuli (bottom).

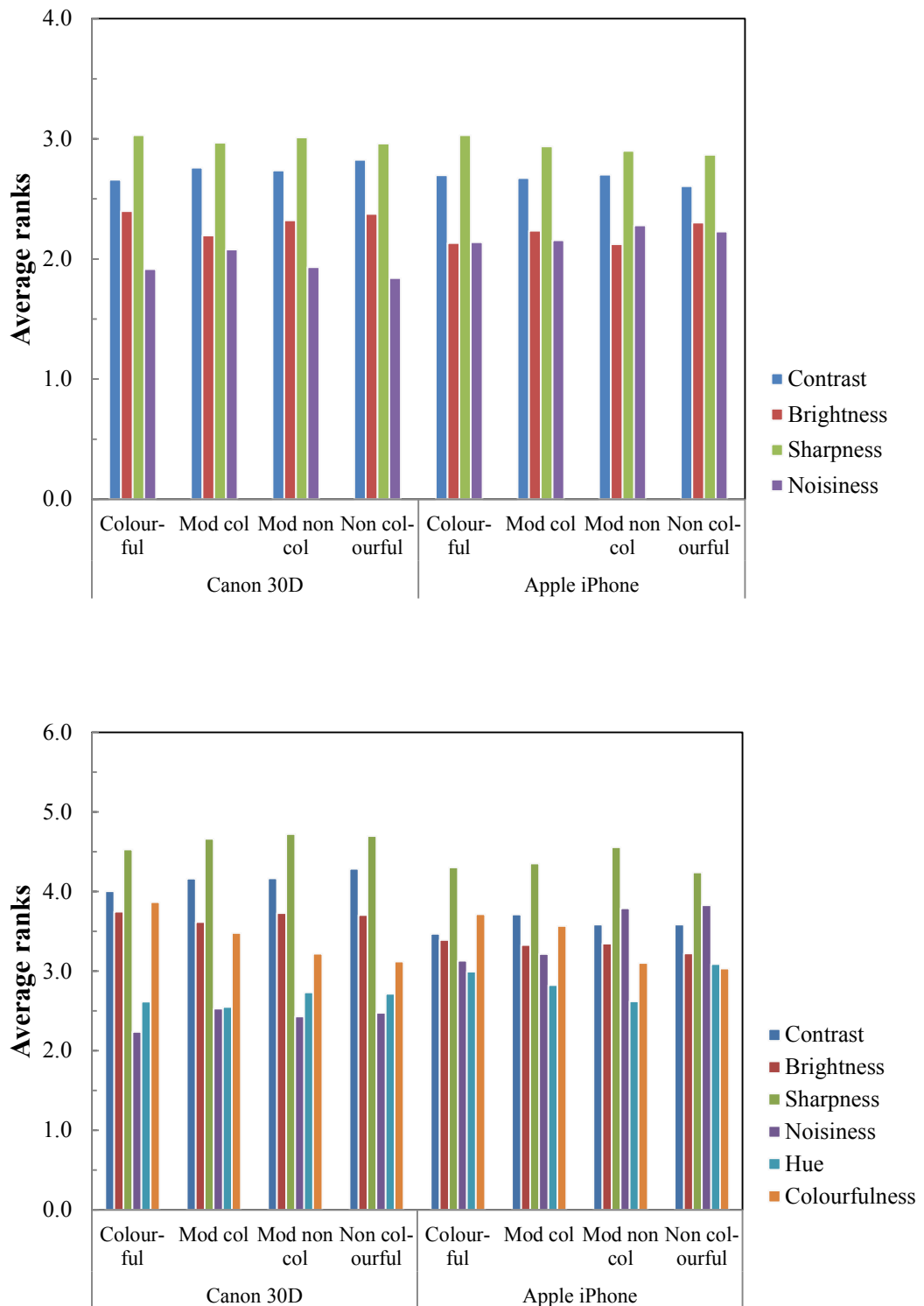


Figure 4-4. Average ranks of the image attributes of test stimuli categorised by their colourfulness. Results from the achromatic versions of stimuli (top) and from the chromatic versions of stimuli (bottom).

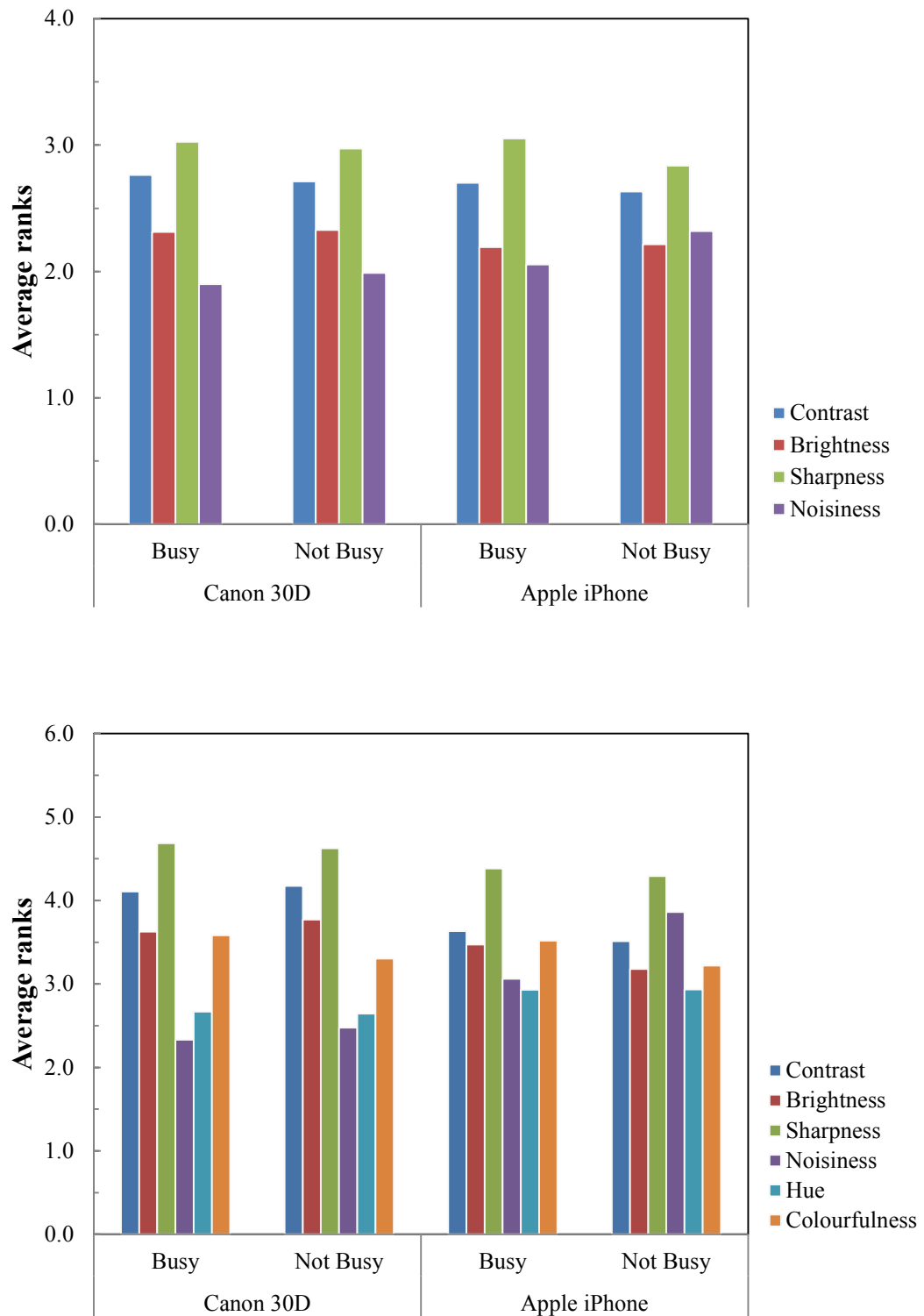


Figure 4-5. Average ranks of the image attributes of test stimuli categorised by their busyness. Results from the achromatic versions of stimuli (top) and from the chromatic versions of stimuli (bottom).

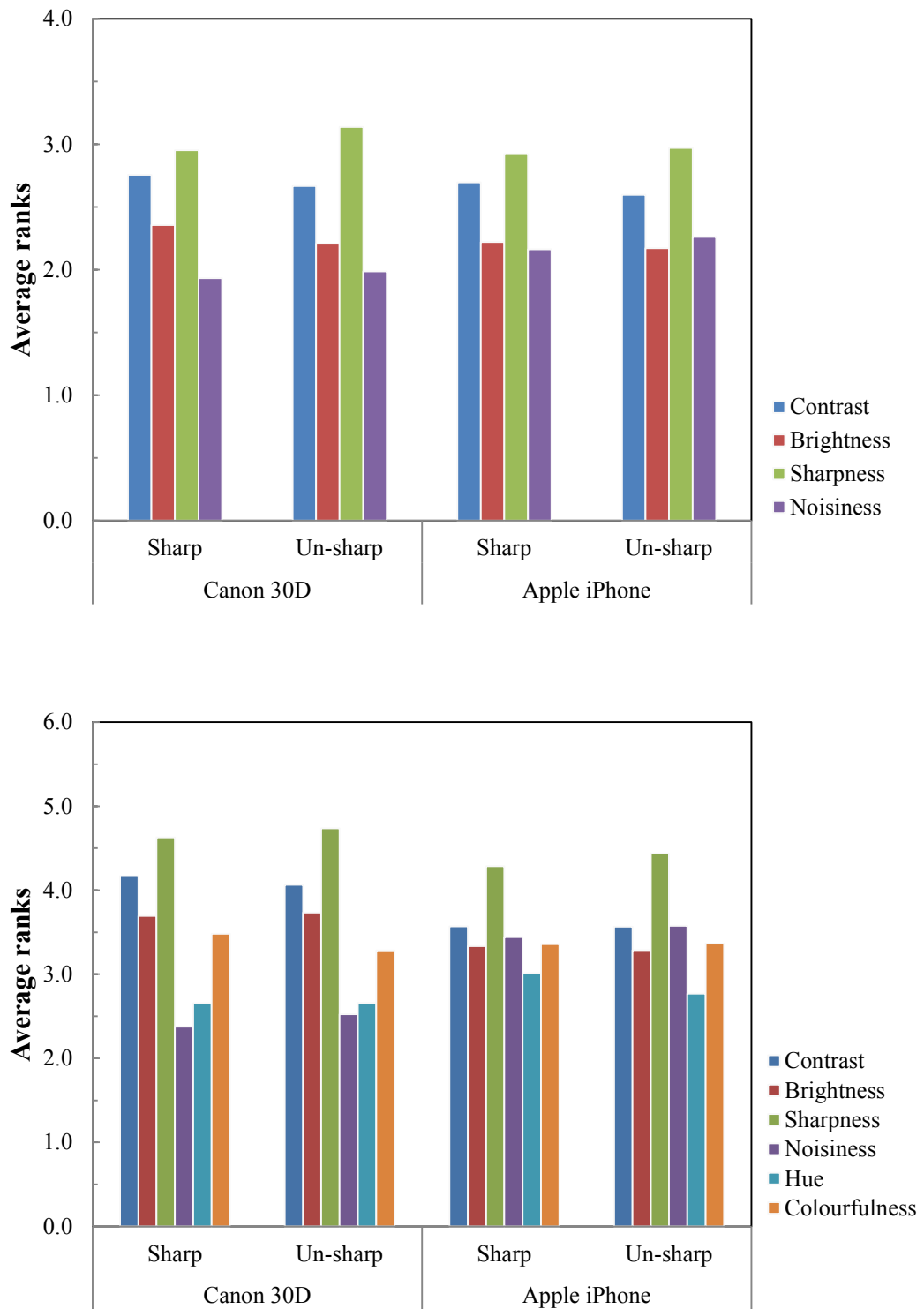


Figure 4-6. Average ranks of the image attributes of test stimuli categorised by their sharpness. Results from the achromatic versions of stimuli (top) and from the chromatic versions of stimuli (bottom).

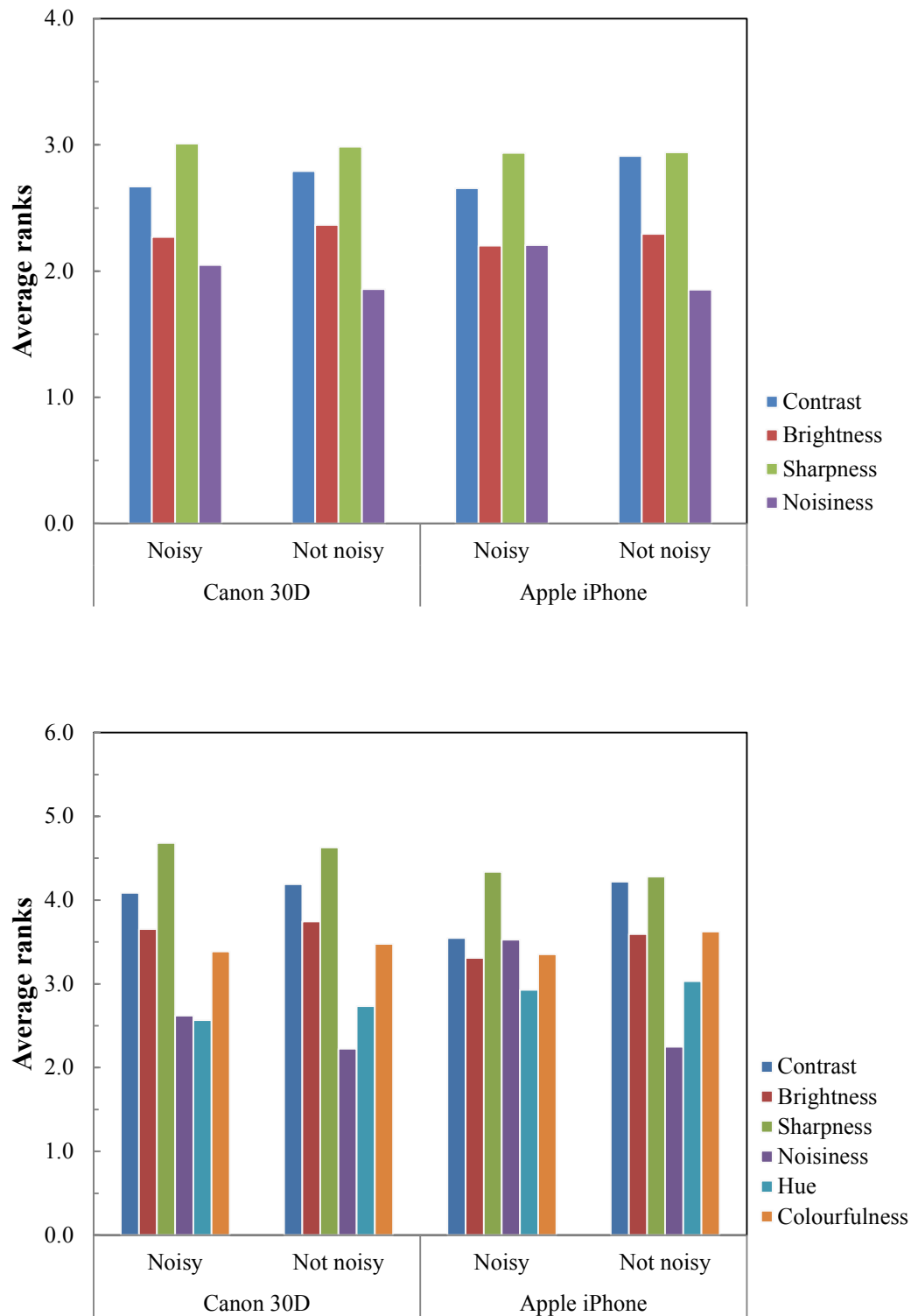


Figure 4-7. Average ranks of the image attributes of test stimuli categorised by their noise level. Results from the achromatic versions of stimuli (top) and from the chromatic versions of stimuli (bottom).

In general, for test stimuli categorised as ‘dark’ and containing higher levels of noise, sharpness was affected less and noisiness was affected more compared with other stimuli classified in different categories (Figure 4-3). For test stimuli obtained by the Canon 30D, the average ranks for brightness was slightly increased as the average lightness of the test stimuli increased (Figure 4-3). In a similar fashion, the average rank of colourfulness was higher for ‘colourful’ stimuli and lower for ‘non colourful’ stimuli (Figure 4-4).

The average rank of noisiness was lower for stimuli categorised as ‘busy’ and higher for stimuli categorised as ‘non busy’ (Figure 4-5). This is due to noise being masked by the high frequency information in the scene (Keelan, 2002, Triantaphillidou *et al.*, 2007, p.35). The average rank of contrast was lower for stimuli that were categorised as ‘sharp’, while the ranks of sharpness and noisiness were higher (Figure 4-6). The noise level of the test stimuli were also related to colourfulness in the chromatic versions (Figure 4-7). The average rank of noisiness was affected more for stimuli categorised as ‘noisy’, indicating that when noise is visible in the image then noisiness becomes an important attribute with respect to changes in image size (Figure 4-7).

4.5 Summary

A novel psychophysical experiment was carried out to investigate the changes in image appearance when images were viewed at different image sizes. Two sets of digital capturing devices with different overall quality were used to record sixty four natural test scenes, with varying scene content and under various illumination conditions. Six image attributes in total (four attributes for achromatic versions) were investigated.

Expert observers with imaging and design backgrounds were selected, because of the complexity of understanding the terms used to describe the perceptual attributes by non-expert observers.

Results from the experiments for the achromatic stimuli indicated that the perceived sharpness and contrast were two most affected image attributes when images are displayed in different sizes, followed by brightness. Noisiness was found to be the least affected attribute. However, results from the experiments for the chromatic stimuli were slightly different. Although sharpness and contrast were again the two most affected attributes by the changes in displayed image size, noisiness came third. This result is not surprising since most of the test images obtained by the Apple iPhone camera possessed some perceptible noise to start with (c.f. Table 4-1). In conclusion, noise can potentially be an attribute affected considerably by display image size, but only when the starting level of noise is quite important.

Also, the results varied with scene content and characteristics. The average lightness level of a scene was found to relate to the attributes of sharpness and noisiness, while the colourfulness of a scene was found to relate more to the attribute of colourfulness (Choi *et al.*, 2007b). The busyness of a scene was also found to relate to noisiness, i.e. as busyness increased noisiness became less important due to visual masking (Keelan, 2002, Triantaphillidou *et al.*, 2007). Finally, the perceptual sharpness of a scene was found to be related to the attributes of sharpness and noisiness (they are complimentary (Johnson and Fairchild, 2000)) whilst the noise level of a scene was related to noisiness.

Although the image size is of one of the important factors affecting the image appearance, there are various other factors should be considered. Due to difficulties in

controlling the ambient lightings in the laboratory, the visual investigations were carried out in a dark environment, to avoid veiling glare. However, this is not the reference viewing condition for sRGB setting. The visual image quality of the test stimuli was nevertheless not affected. Furthermore, the CG210 display used in this experimental work was found to exhibit inaccurate tone reproduction and considerable positional non-uniformity characteristics (c.f. Section 3.2), which may have influenced the results.

Chapter 5

Psychophysical investigation 2: Evaluation of changes in perceived sharpness with changes in displayed image size

This chapter is concerned with the quantification of the degree of change in perceived image sharpness with respect to changes in displayed image size. This was achieved by collecting data from the visual sharpness matching investigations using the method of adjustment in a dark viewing environment. This chapter first describes a method adapted from ISO 20462-3 (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005), employed to create a series of frequency domain filters for sharpening and blurring. The filters are designed to provide equal intervals in image quality from a certain viewing distance. The effect of bi-cubic interpolation on image sharpness is also examined. Secondly, it explains the method and steps used for the evaluation of changes in image quality due to the changes

in displayed image size. Finally, the validation of results obtained from the sharpness matching experiments, and the details of calibration of the relative quality scale to sharpness JND scale are explained.

5.1 Preparation of test stimuli

5.1.1 System tone reproduction

The experimental methodology involved the measurement of the transfer functions of the capturing and display devices and the consequent determination of the combined (overall) system gamma, which was further used for signal linearisation during SFR measurements. Measurement of the transfer function of the Canon 30D camera and EIZO ColorEdge CG245W display systems used in this investigation were carried out individually and described in Chapter 3. The combined transfer function was also measured and used in this investigation for both simplicity and accuracy.

The LCD display device was calibrated to a white point luminance of 120cd/m^2 , a gamma of 2.2, and a colour temperature of D_{65} . For the measurement of the combined (camera-display) transfer function, the average pixel value of the patches of captured greyscales presented in Section 3.1 was displayed on the calibrated display. The output luminance values of the displayed patches were measured using a calibrated Konica Minolta CS-200. Normalised output luminance values were plotted against normalised input luminance values in log-log scale, as shown in Figure 5.1. The overall gamma of $\gamma=1.21$ was derived from a linear portion of the curve (c.f. Section 3.1.3).

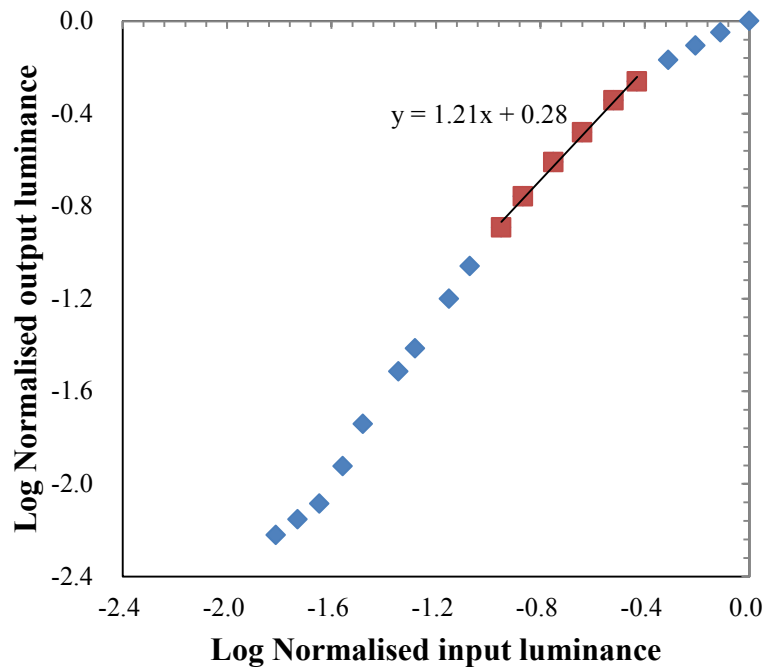


Figure 5-1. Transfer function of the Camera-Display combined system.

5.1.2 System SFR

The slanted edge method (BS ISO 12233:2000 (*Photography--Electronic still picture cameras--Resolution measurements*, 2000)) was implemented to measure the combined spatial frequency response (SFR) of the capturing and display devices. A test target with scalable vector graphics (SVG) patterns containing vertical and horizontal edges was created using Imatest software (Imatest, 2013). The test target had a contrast ratio of 3:1, at a gamma of 2.2. It was displayed on the calibrated EIZO LCD employed in the sharpness matching and was captured using the Canon EOS 30D camera with the zoom lens set to focal length of 22mm, from a distance of 100cm. This distance was chosen to avoid aliasing originating from the LCD pixel structure during the image acquisition, while maintaining the spatial frequencies of interest.

The system gamma of $\gamma=1.21$ (c.f. Section 5.1.1) was taken into account for the gamma correction in the computation of the combined system SFR. The horizontal and

vertical SFRs were weighted 1/3 and 2/3 to determine the ‘average’ SFR of the combined system at each of the aperture stops (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005).

The spatial frequency units were then converted from cycles/pixel to cycles/degree.

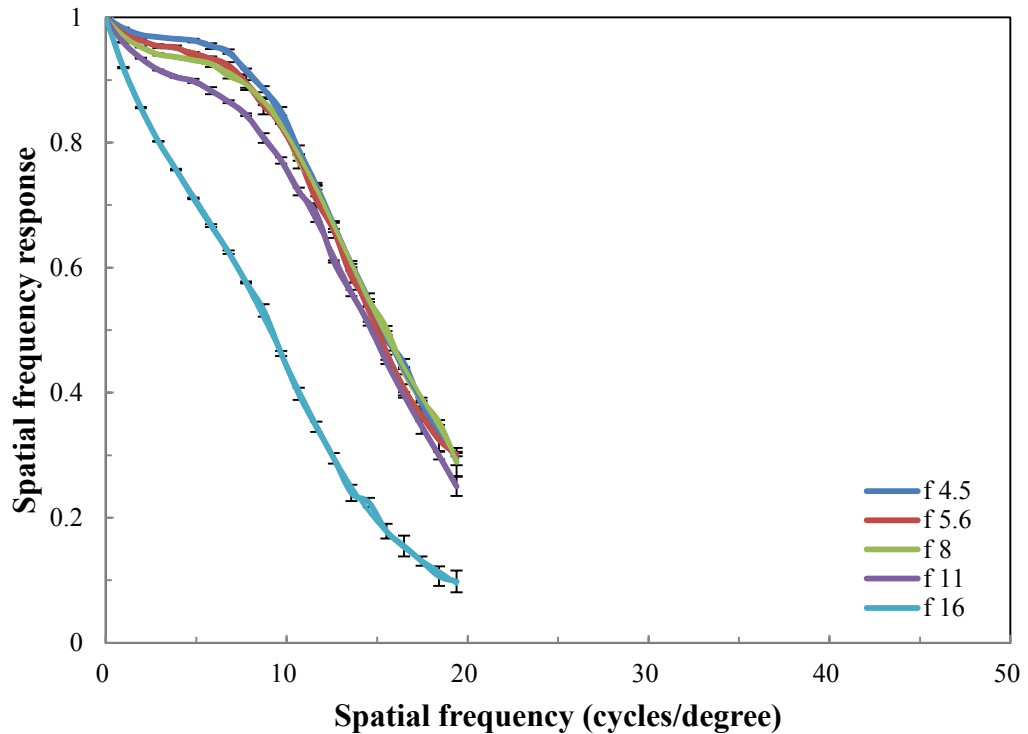


Figure 5-2. Spatial frequency responses (SFRs) of the combined system at major aperture stops.

Three exposures were made for the SFR measurements at various lens apertures by varying ISO settings (ISO 100, 800, and 1600). In addition to the computation of SFRs, average standard error of mean (SEM) were also calculated. The SFRs obtained with the apertures at $f11$ and larger were found very similar, while the SFR with aperture at $f16$ was much lower, as shown in Figure 5-2. All scenes used in this investigation were captured with apertures at $f16$ or larger. Therefore corresponding k values for aperture at $f16$ and $f8$ were selected as reciprocal measures of the system

bandwidth. Note that k values are key parameters in the determination of the sharpening and blurring filters (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005, p.11) and are discussed in the next section.

5.1.3 Determination of the reciprocal measure of the system bandwidth, k

A set of model curves was computed to determine the reciprocal measures of the combined systems, using Equation 2.16 by varying the values of k . They are shown in Figure 5-3. The combined system SFRs (for the set observation distance) were compared with the modelled curves at modulation of 0.5 and 0.3. Then the k values of the nearest modelled curves were selected. The k values, which represent the ‘shape parameter’ of the model curve, were $k=0.030$ for apertures up to $f11$ and $k=0.047$ for $f16$. The secondary standard quality scale, SQS_2 , values associated with the system bandwidth, k , at each aperture were then calculated using Equation 2.17 (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2012). The SQS_2 values associated with $k=0.030$ and $k=0.047$ were found to be approximately 27 and 20, respectively, as shown in Figure 5-4.

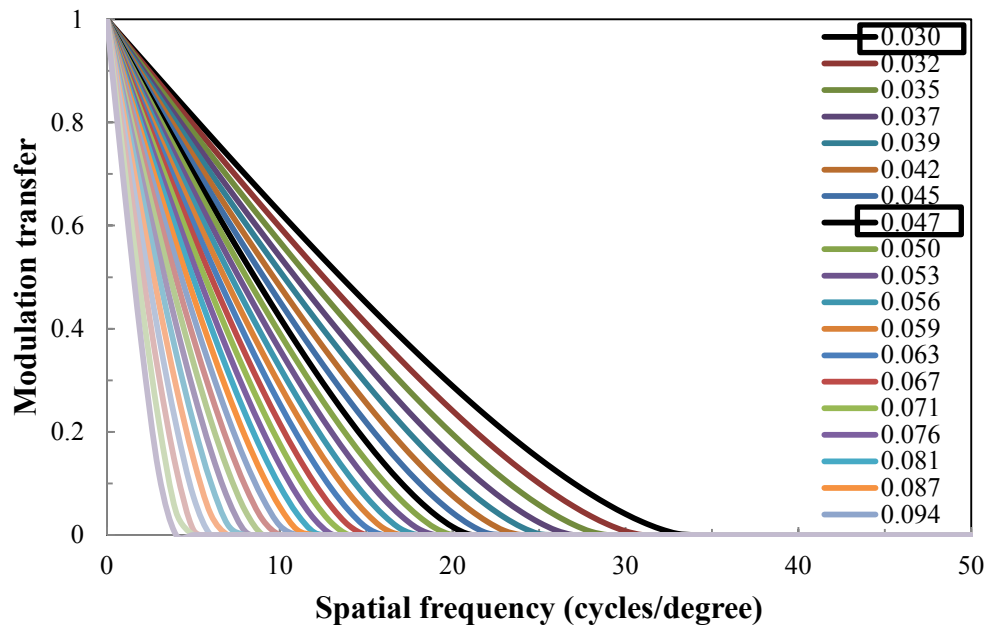


Figure 5-3. Modelled MTF curves with the various k values.

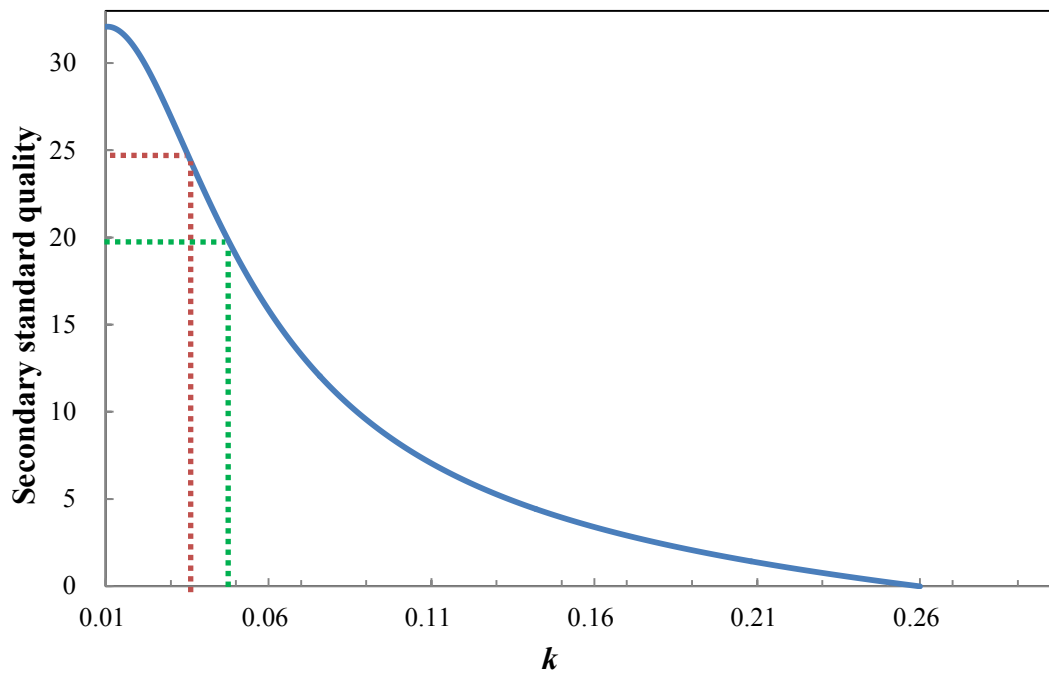


Figure 5-4. Secondary standard quality value at $k=0.030$ and $k=0.047$.

5.1.4 Sharpness filters

A total of thirteen modelled curves with k values associated with SQS_2 values, ranging between 27 and 14 for the aperture at $f/11$ and below and between 20 and 8 for $f/16$, were selected with constant intervals of 1 SQS_2 . The curves were then divided by the corresponding combined system SFR. SFRs beyond the Nyquist frequency, λ_N , of the LCD display for the given observation distance were replaced by 0. The resulting curves represented the functions to be used for blurring; they are illustrated in Figure 5-5. Vertically flipped versions of these represented the functions for sharpening, shown in Figure 5-6. The x -axes in the Figures 5-5 and 5-6 represent the image dimensions in pixels, thus leading to the filter functions is dependent upon the image size. The 1-D Gaussian filter functions (Gonzalez and Woods, 2002, p.175) in Figures 5-5 and 5-6 were obtained for the dimensions of the original images, using Oakdale Engineering Datafit software v9.0 (Oakdale Engineering, 2007).

Equation 5-1 represents the 2-D circularly symmetric Gaussian filter functions (Easton, 2010, p.197) for blurring and sharpening, H ,

$$\begin{aligned} H &= 1 - e^{-a \times b^D} && \text{for blurring} \\ H &= 1 + e^{-a \times b^D} && \text{for sharpening} \end{aligned} \quad (5-1)$$

where D is the digital image dimensions and their spectra, and a , and b are the variables which controls widths of the filter apertures.

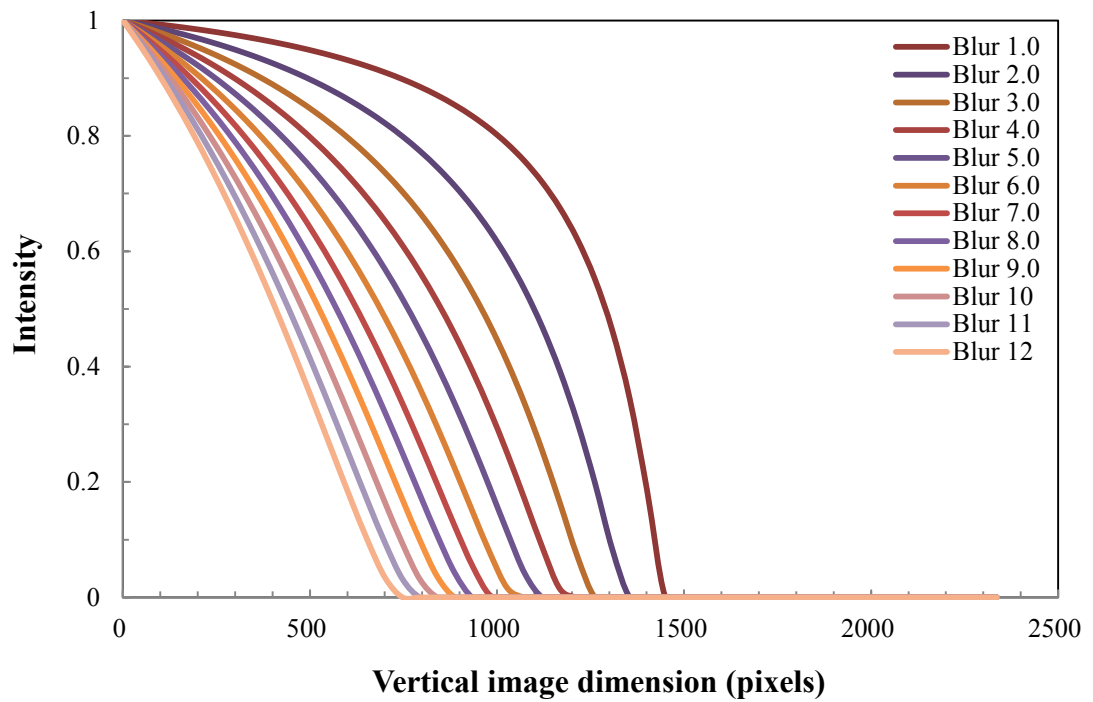


Figure 5-5. Cross section of blurring filters for the images taken at $f/11$ and below.

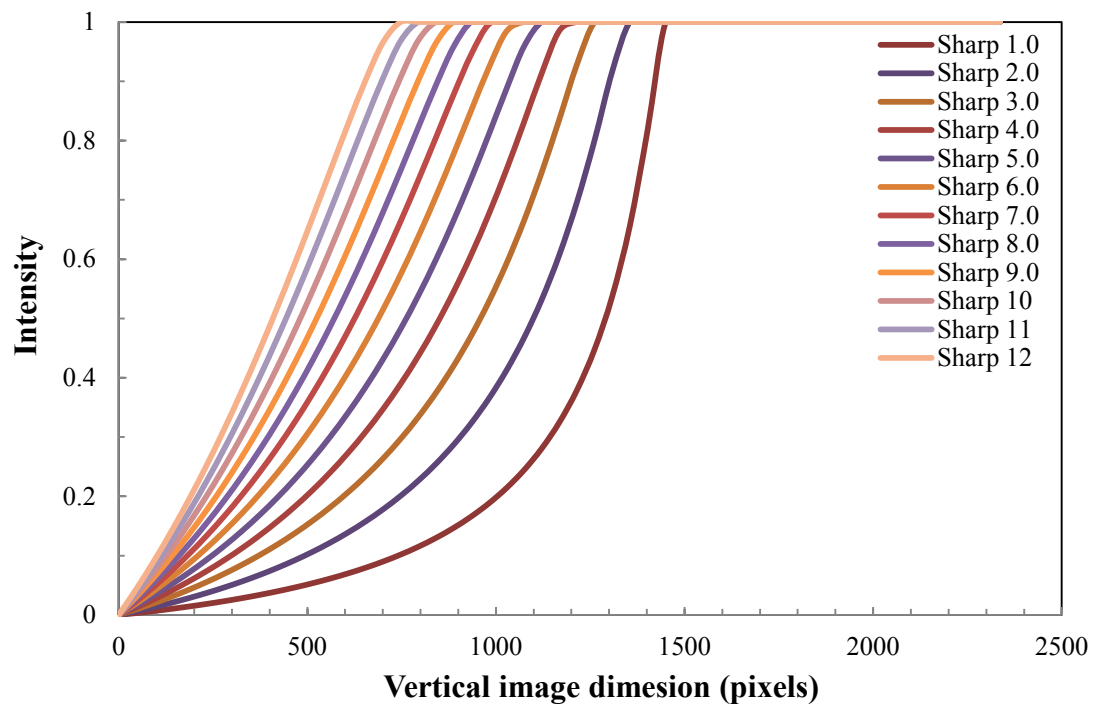


Figure 5-6. Cross section of sharpening filters for the images taken at $f/11$ and below.

5.1.5 Frequency domain filtering and bi-cubic interpolation

The filtering operations were carried out using MATLAB. The original images were first converted from the sRGB space to the linear with luminance $R'G'B'$ space. The filters were applied to the spectra of the R, G, and B channels. The mean pixel values of the images were subtracted before the filtering and added back after the filtering to maintain the mean luminance of the scenes unaltered.

A total of 25 ruler images, possessing different image quality levels with intervals of 1 SQS₂ (original, 12 blurred versions and 12 sharpened versions) were generated for each original size image. We discuss the relationship between 1 JND in perceived sharpness and 1 SQS₂ in image quality in a later section (c.f. Section 5.3.3). The ruler images were then resized to five versions of the same scenes of different sizes by bi-cubic interpolation to minimise interpolation artefacts with a cost of sharpness loss (Park *et al.*, 2012). The test image dimensions were 744(H)×560(V) pixels, 635(H)×478(V), 526(H)×396(V), 449(H)×338(V), and 372(H)×280(V) and represented large, large-medium, medium, medium-small and small sizes commonly displayed on computer and mobile device monitors. The smallest size was based on prevalent dimensions of LCD of digital SLR cameras. The largest version was approximately half the size of CG245W LCD's native horizontal and vertical resolutions.

5.1.6 Effect of bi-cubic interpolation on image quality

The test stimuli were converted from the original capture size to five different sizes using bi-cubic interpolation. Because the interpolation affects the spatial characteristics of the images (Jin *et al.*, 2009), effect of the bi-cubic interpolation on image quality was examined.

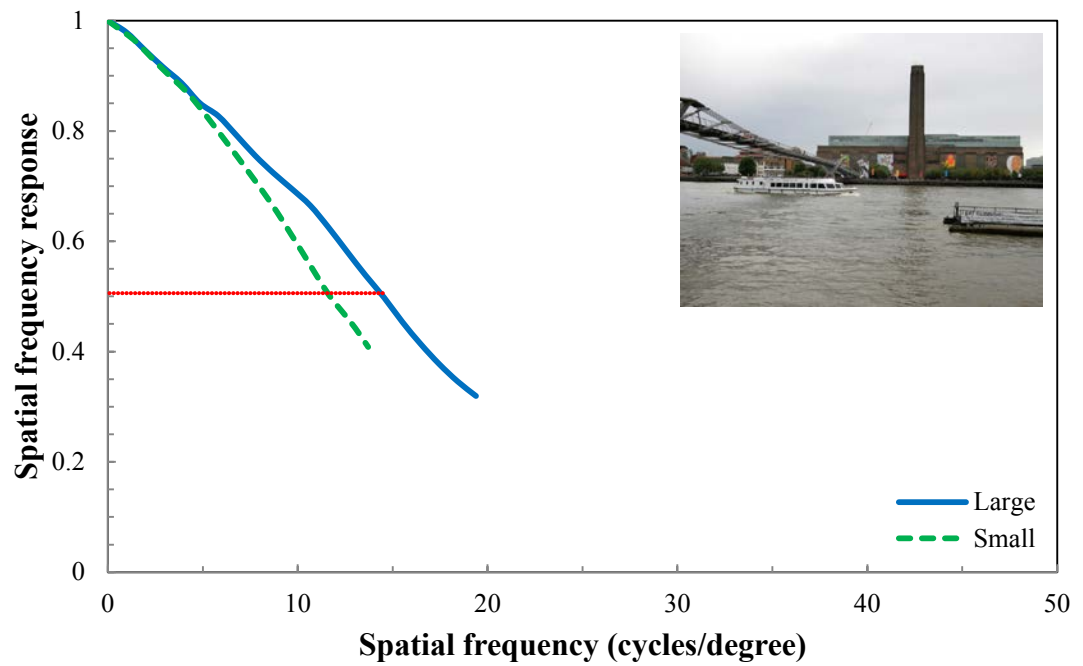


Figure 5-7. Effect of the bi-cubic interpolation on SFR, Tate Modern scene.

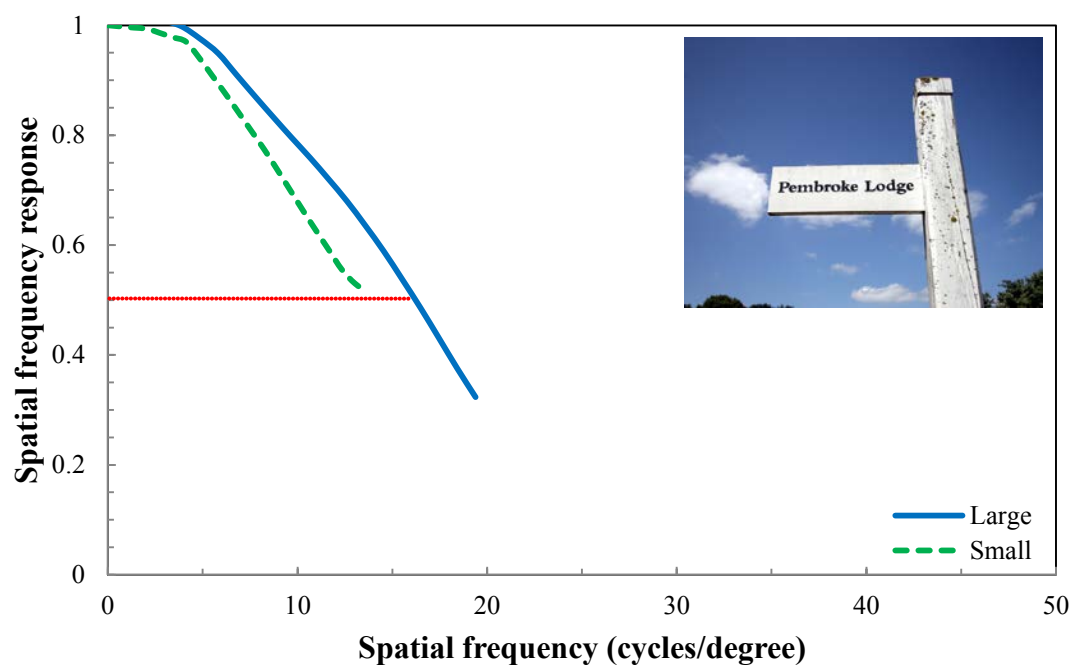


Figure 5-8. Effect of the bi-cubic interpolation on SFR, Pembroke lodge sign scene.

To quantify the effects of the resizing operation on image quality, two scenes possessing strong, measurable edges were selected by visual inspection. The SFRs of the two different scenes were measured for large size versions and small ones from a selected edge in each scene.

Image quality loss, as seen in Figures 5-7 and 5-8, was detected after the interpolation. The extent of this effect was however observed to be scene independent. The less sharp image, shown in Figure 5-7 to have a lower SFR, was affected approximately 3 SQS₂ at the SFR of 0.5 in standard quality scale by the interpolations, while the sharper image, with a higher SFR shown in Figure 5-8, was affected approximately 2 SQS₂ at the SFR of 0.5. The effect of the bi-cubic interpolation was taken into account for the data analysis in Section 5.3.

5.2 Psychophysical investigation

5.2.1 Display settings and calibration

The EIZO ColorEdge CG245W 24.1'' LCD, driven by a Dell Optiplex 760 computer with an ATI Radeon HD 3450 graphics controller, was used in the psychophysical investigation. The LCD has a native spatial resolution of 1,920×1,200 pixels and a tonal resolution of 24bits (with a DVI connector). The system was set to a white point luminance of 120cd/m², a gamma of 2.2 and a colour temperature of D₆₅, using the GretagMacbeth Eye-One Pro with Profilemaker v5.0. Daily calibration was carried out using the built-in calibration sensor throughout the period of the psychophysical investigations.

5.2.2 Software preparation and interface design

The application employed in the sharpness matching experiment was written in PHP, HTML and CSS, the user interface being controlled using JavaScript. It was tested and optimized for Mozilla Firefox v5.0 web browser (Mozilla, 2013). A mid-grey background in luminance (pixel value of R=G=B=186, at a gamma of 2.2) was selected. The application gathered some personal information provided by the observers before the experiments started and it automatically wrote the observation data and saved them in a comma-separated value (CSV) file. The display interface is illustrated in Figure 5-9.

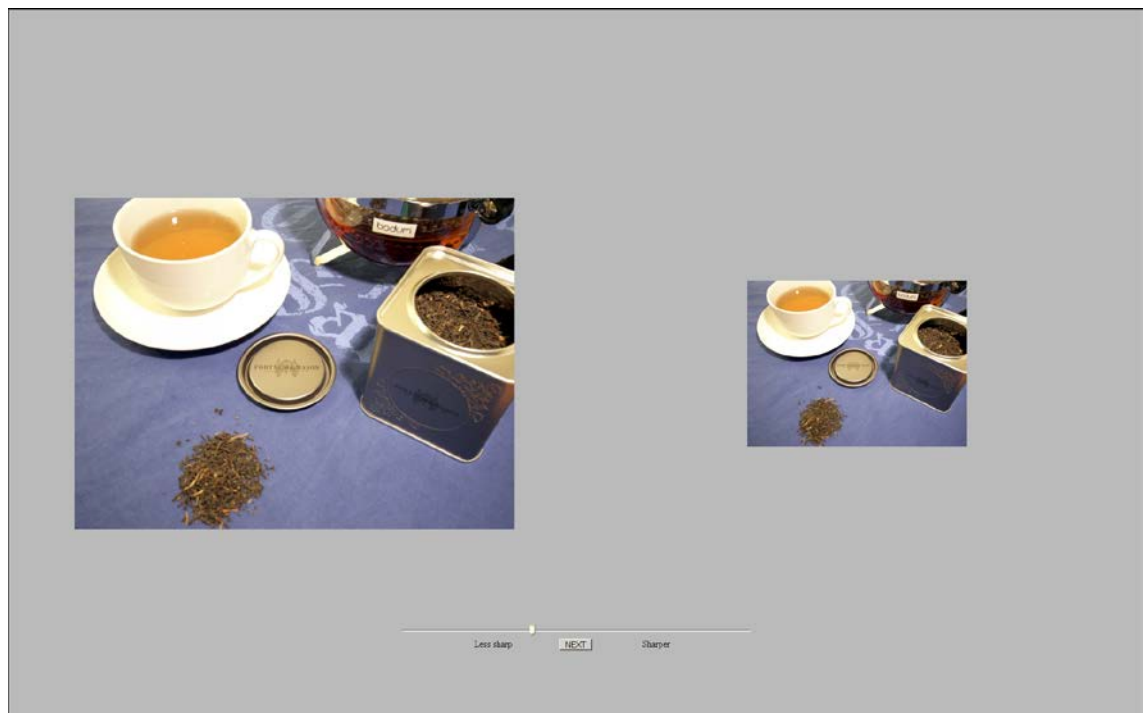


Figure 5-9. Display interface of sharpness matching test with a slider.

5.2.3 Sharpness matching experiment

Visual sharpness matching experiments, using a slider controlled by the computer mouse, were conducted in a totally dark environment as described in Section 4.2.3. Observers were seated on a comfortable seat with a chin rest to hold the observation

distance at 60cm from the display and were requested to only move their eyes from side to side. During the tests, a randomly selected test image was displayed simultaneously at two different sizes. The test images were displayed with random display position, one on the left side and the other on the right side of the display, to minimise the impact caused by the positional non-uniformity and viewing angle dependency of the display device (c.f. Section 3.2.9). Observers were asked to match the sharpness of the smaller ‘test’ images to that of the larger ‘standard’ images using a slider. The slider was programmed to simulate to the user an enhancement of quality of the images in response to changes in the slider position, by replacing the test image with the appropriate ruler image, according to the selected slider position.

Preliminary experiments consisting of three sharpness matching sessions was carried out using large, medium, and small size test image sets. The purpose of this step was to select ‘average’ scenes and consistent observers. A total of twenty-two observers, 10 females and 12 males, participated in the experiment using all sixty-four scenes. Their ages ranged between 20 and 40 years old; fifteen observers had imaging and design backgrounds (considered as experts). Each observation took less than one hour per session; one session per day was conducted to avoid fatigue.

As a first step for the elimination of extreme observers, the responses obtained by each individual observer were summed and the observers who responded in opposite direction to the majority of observers were discounted. Then the results obtained from all the remaining observers were averaged for each individual scene and the standard deviation, σ , was computed. As the purpose of this step is to select average scenes and consistent observers, to eliminate extreme results for each scene, any observations outside mean $\pm 1\sigma$ range (68% confidence interval) were excluded. After the elimination

of the extreme observations, means and standard deviations, σ , were re-calculated for each scene and for each observer. The above calculations were repeated using the median value $\pm 1\sigma$. Sixteen common images and seven average observers, 3 females and 4 males, from the above steps were selected for the final experiments with the images at five different sizes.

Final sharpness matching experiments at five different image sizes were carried out using 16 ‘average’ scenes by seven ‘average’ observers selected by the step above. The experiments consisting four sharpness matching sessions: small size to large size, medium-small size to large size, medium size to large size, and large-medium size to large size. Each observation took less than 20 minutes per session, with sufficient time being allowed between experiments for recovery from possible fatigue.

5.3 Results and discussion

For the analysis of the responses from the psychophysical experiments, the mean, μ , changes in image quality and standard error of the mean, SEM, was calculated for each scene and size pair. During the calculation, the effect of bi-cubic interpolation on image quality was taken into account (c.f. Section 5.1.6).

5.3.1 Results from the psychophysical tests

Observations from matching the sharpness of the small version image to that of the large version image, resulted in an average image quality loss of 9.179 JNDs (in secondary standard quality scale, SQS₂), with an average standard error of mean (SEM) of 1.151. The range of image quality losses was from 7.79 to 10.50. The losses in image for each scene, along with standard error, are plotted in Figure 5-10.

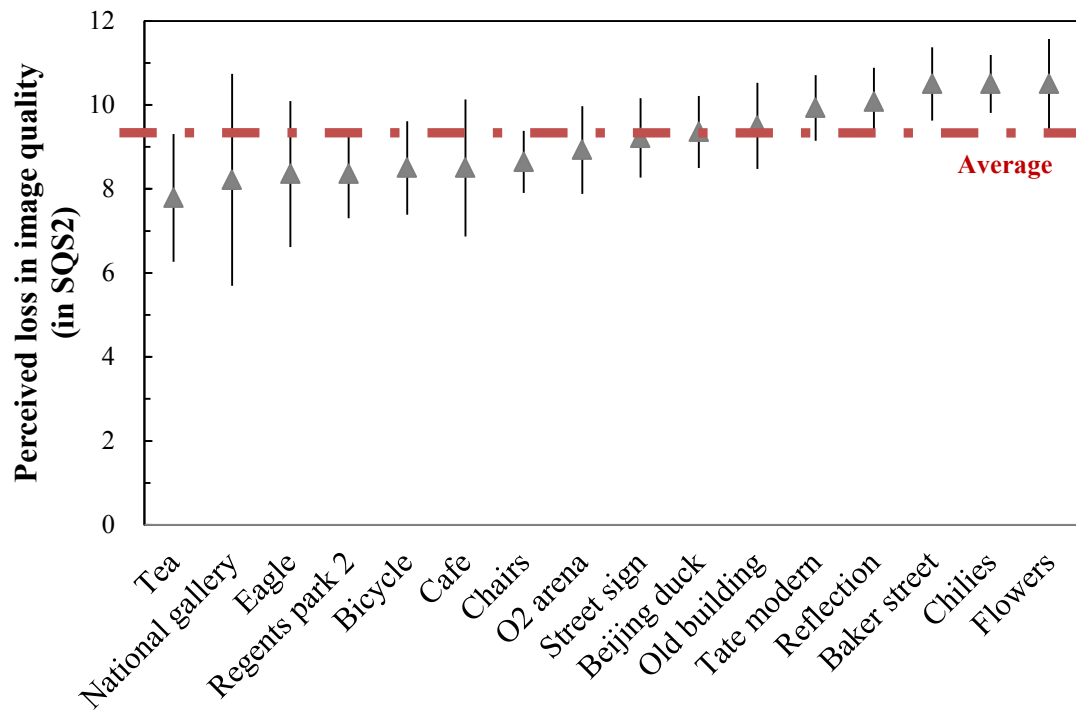


Figure 5-10. Average perceived loss in image quality from the small vs. large experiment for each scene with SEM.

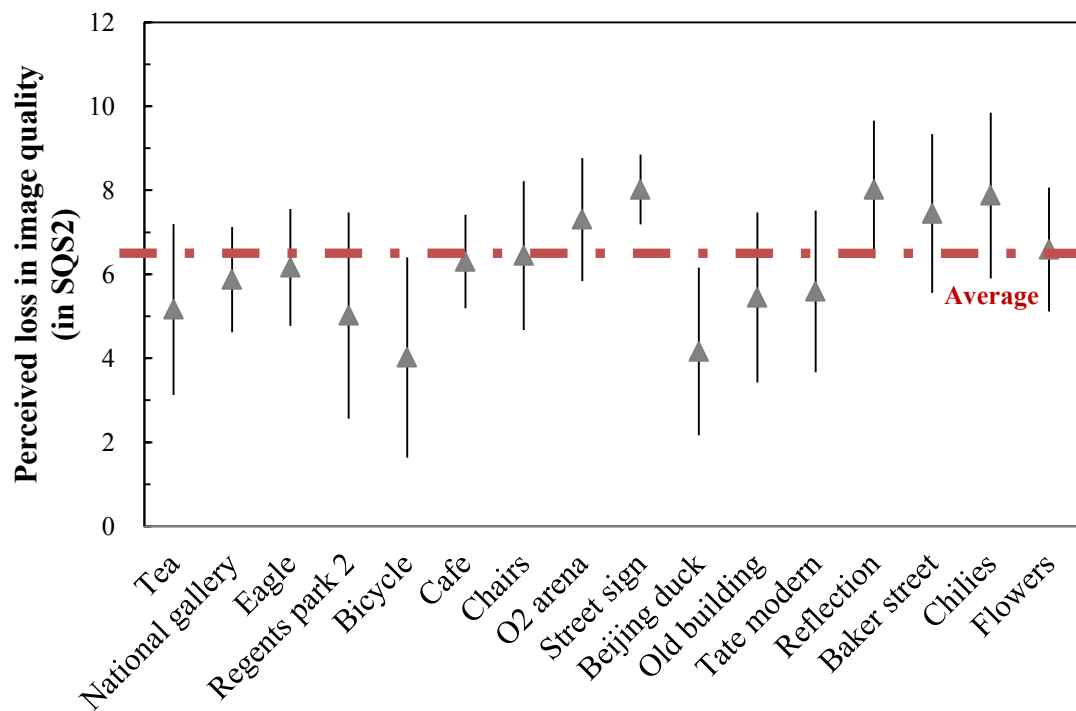


Figure 5-11. Average perceived loss in image quality from the medium-small vs. large experiment for each scene with SEM.

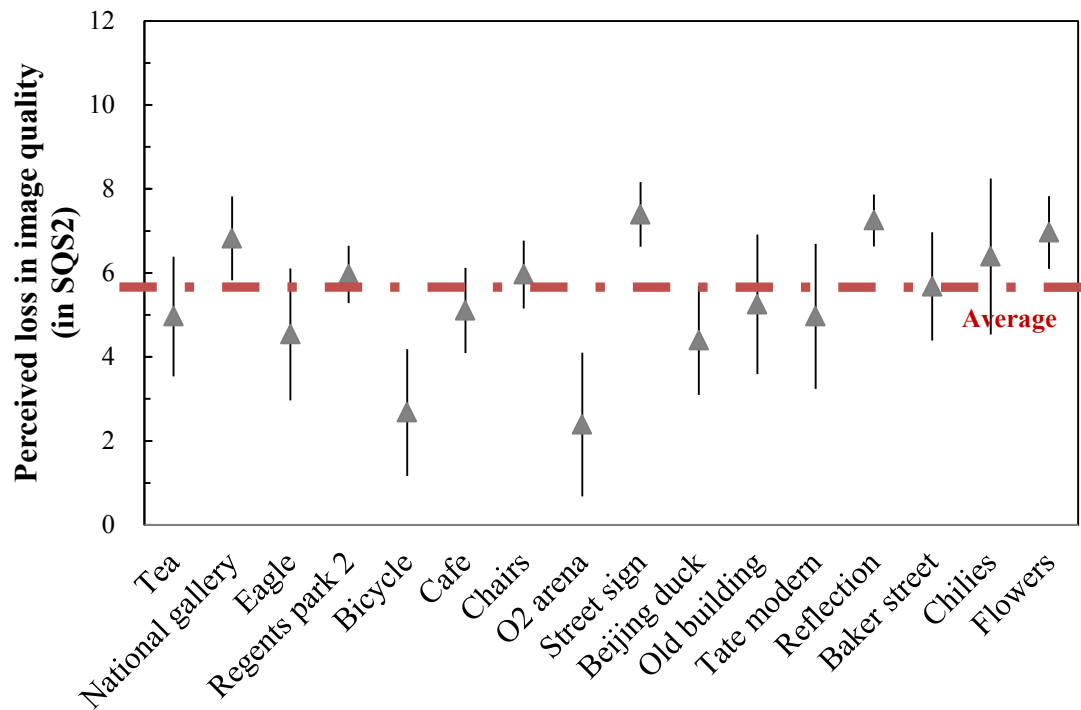


Figure 5-12. Average perceived loss in image quality from the medium vs. large experiment for each scene with SEM.

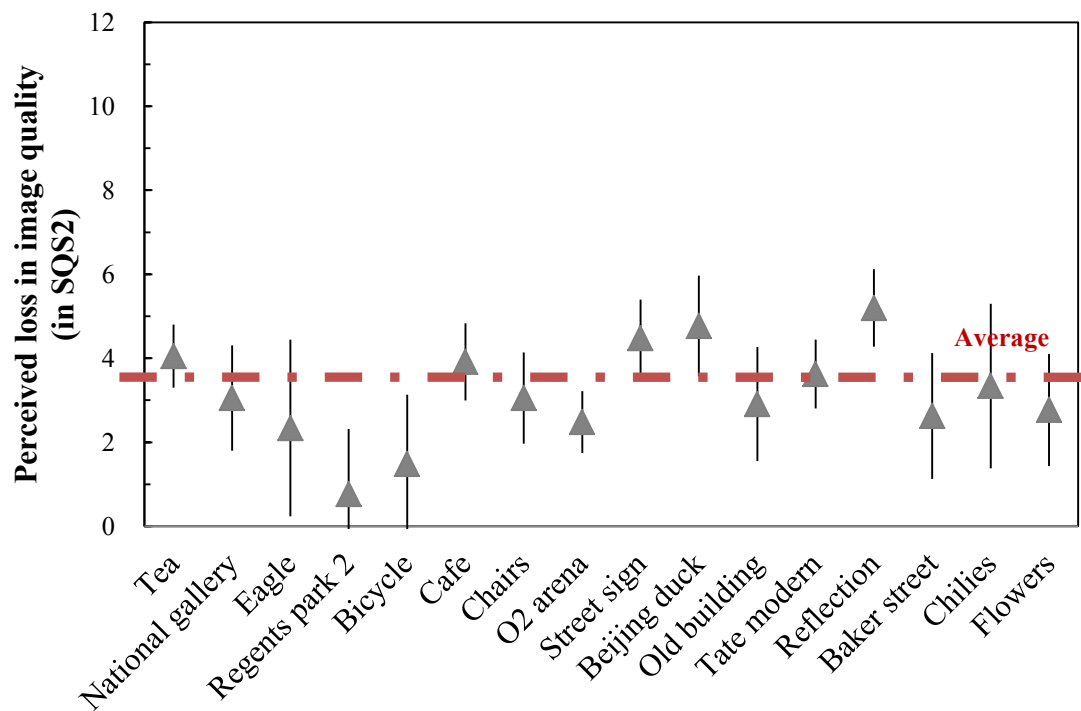


Figure 5-13. Average perceived loss in image quality from the large-medium vs. large experiment for each scene with SEM.

From the experiment of the medium-small against large pairs of images, an average change was 6.214 JNDs with an average SEM of 1.726. The range of quality losses was from 4.02 to 8.02. From the medium version against the large version image, an average change was 5.420 JNDs with an average SEM of 1.237. The range of quality losses was from 2.39 to 7.39. And the large-medium version against the large version matching experiment showed that the average image quality change was 3.179 JNDs with an average SEM of 1.253. The range of quality losses was from 0.77 to 5.20. The results are plotted in Figures 5-11 to 5-13.

In addition to the above figures, the average changes in perceived image quality in SQS_2 from all four experiments were plotted as a function of displayed image size in Figure 5-14. The figure clearly shows that the perceived sharpness was affected by changes in the displayed image size linearly. Smaller version images were perceived sharper than that of the larger version whilst the relationship between perceived sharpness and image size was very close to an inverse linear relationship. Mirrored data at zero point have also been estimated by extrapolation and plotted as linear function to predict change in perceived sharpness when images may be displayed at larger scales. This assumes that the relationship remains linear. The linear trend line showed the relationship as; $y = -0.11x - 0.05$.

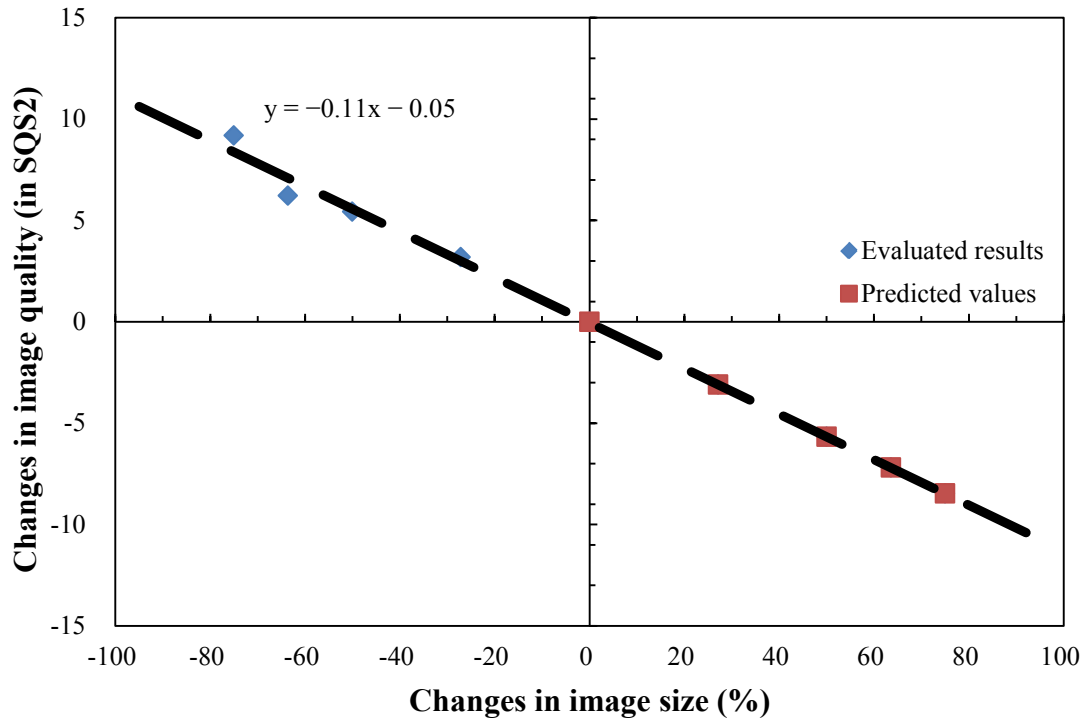


Figure 5-14. Perceived changes in image quality with respect to the changes in displayed image size (blue) and estimated changes (red) in non-calibrated relative image quality JND scale (SQS₂).

5.3.2 Validation of the results

In order to validate the results acquired by the sharpness matching investigations, a series of pair rating experiments was conducted by adapting the magnitude estimation method. A total of sixteen large size average scenes (c.f. Section 5.2.3) and their corresponding smaller versions, one unmodified and one sharpness modified version, were prepared. A total of 7 average participants, who carried out the sharpness matching investigation, carried out the experiments under the same experimental environment described in the earlier section (c.f. Section 4.2). A randomly selected test image pair was displayed to the observers at a time during the experiments. The observers were then asked to rate the test image pair in terms of their appearance matching (from 10 being the most matching to 1 being the least matching). Because the experiments were conducted without a reference, the observers were asked to avoid the maximum and

minimum rates for the first image pair. Observer calibration was carried out for the analysis of data.

Results from these experiments confirmed that the image pairs with sharpness modified version, average rating of 5.00, appeared to be the better matching compared with those with unmodified version, 4.62, as shown in Figure 5-15.

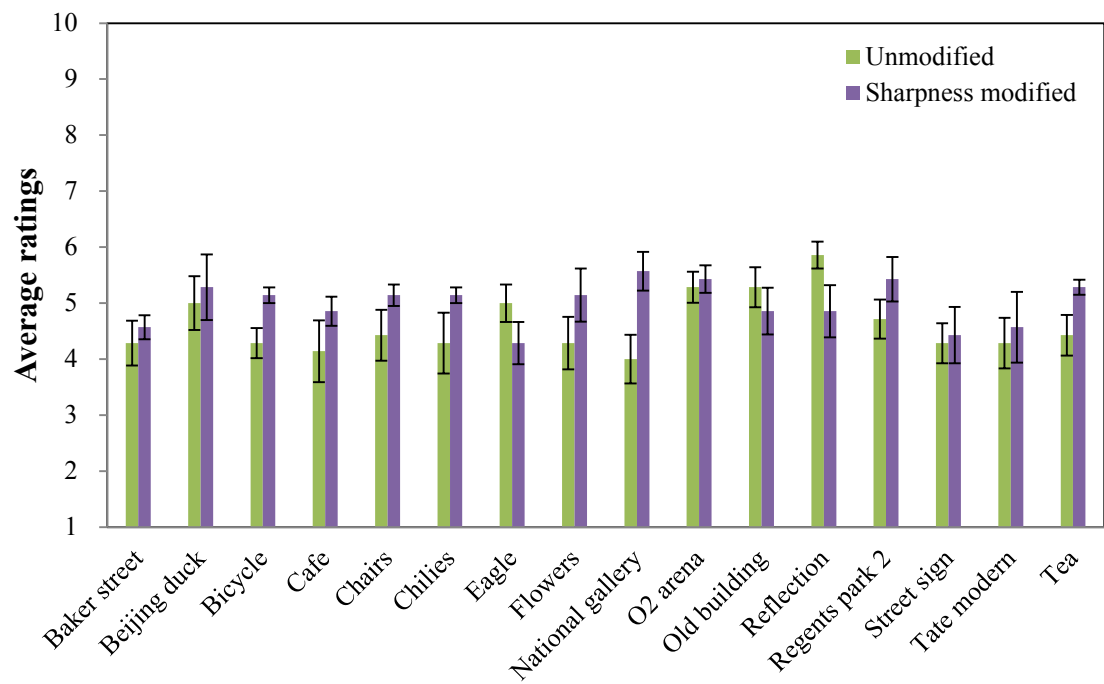


Figure 5-15. Average ratings of the unmodified pairs and the sharpness matched pairs.

5.3.3 Evaluation of step interval and calibration of changes in sharpness

JND scales

A series of paired comparison experiments to evaluate the sharpness step intervals were conducted using all sixty four scenes. For the step interval evaluation, a new set of filters with smaller intervals (half of the intervals used for sharpness matching) was created for sharpness enhancements to increase accuracy. In order to design the

experiments more efficient (c.f. Section 2.3.3.2), central region of the scale (original ± 6 steps) was used.

Three male expert observers participated to a total of one hundred and ninety two sessions. Each observation took less than 10 minutes per session and a maximum of 10 sessions per day was conducted to avoid fatigue.

The outcome was used to calibrate the results plotted in Figure 5-14 into perceptually meaningful JND scale for the sharpness attribute. From the experiments for the sharpness step validation, an average of 0.71 JNDs (in SQS₂ scale) was found to be 1 JND in perceived sharpness. The results obtained in Section 5.3.1 were then calibrated and plotted in Figure 5-16. The linear trend line showed the relationship as; $y = -0.159x - 0.0695$. The change in perceived sharpness was as much as 11.86 JNDs with 75% change in the displayed image size.

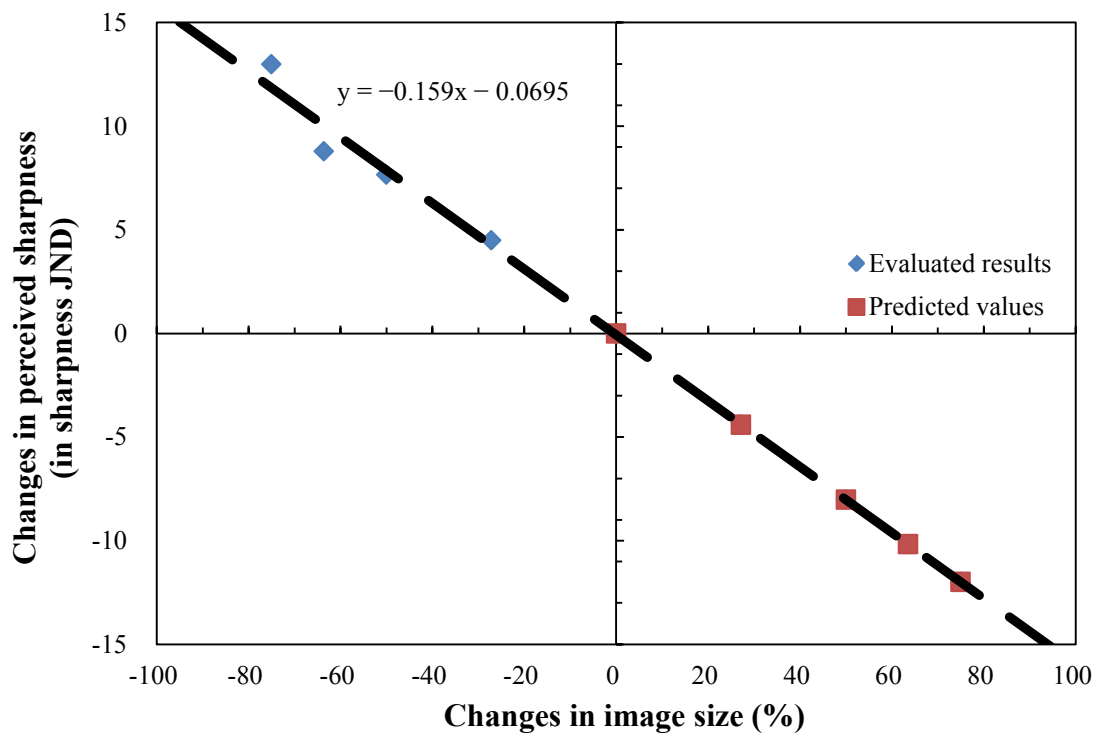


Figure 5-16. Changes in perceived sharpness with respect to the changes in displayed image size (blue) and estimated changes (red) in sharpness JND scale.

5.4 Summary

From the previous experiments, sharpness attribute was identified as the most affected perceptual image attribute by changes in displayed image size. Therefore, a series of five psychophysical experiments were carried out to quantify the changes in perceived sharpness with respect to the changes in displayed image size using the method of adjustment. This method was chosen because it allowed direct evaluation of the visual differences between a pair of images of different sizes, viewed at the same time and on the same display. A total of sixty-four natural scenes, captured using Canon 30D camera, with varying scene content were initially selected. For the image display, a new CG245W monitor, exhibiting better overall characteristics was employed. This was because of the monitor used in the rank order investigation (c.f. Chapter 4) exhibited positionally inhomogeneous characteristic, which made the monitor unsuitable for the purpose of the matching experiment.

A set of 25 images of varying image sharpness with an equal quality interval were created for each original, using frequency domain filtering by adapting the method described in ISO 20462-3. The filtered images were resized to generate five different sizes: large, large-medium, medium, medium-small and small. The observers were requested to start to match the small version to the large reference since the difference in displayed image size was the biggest. Results from all four psychophysical experiments indicated that the smaller version images were perceived as sharper (i.e. better quality) than the reference ones with approximately linear trend. Average difference of perceived image quality between the reference version images and the small ones was approximately 9.18 JNDs.

The softcopy ruler images were created by an adaptation of ISO 20462-3. In different literature (Keelan, 2002, p.73), Keelan who is one of the co-authors of the standard has described one relative quality JND in Secondary Standard Quality Scale (SQS₂) as a multivariate JND increment, which is larger than a univariate increment. Univariate increments vary in one attribute, such as in sharpness or noisiness only, where multivariate increments are based on all attributes that affect overall quality. In the case of sharpness, Keelan found JND increments of quality to be approximately twice as large as JND increments of sharpness (Keelan, 2002). Keelan calibrated the relative quality JND scale in the standard using images with negligible artefacts and noise, and with excellent colour and tone rendition (Jin *et al.*, 2009).

However, the image set used in this experiment was carefully selected to comprise examples of imagery captured by ordinary camera users, rather than professionals. Therefore, the images were expected to be “not at the best quality the camera system can produce” in colour and tone as well as in sharpness. Thus, the results obtained from the sharpness matching experiments do not in all cases correspond directly to the standard quality scale, SQS₂, presented in the standard. Therefore, calibration of the results from the SQS₂ into perceived sharpness scale was carried out next. Since the results acquired from the matching experiments were in image quality (multivariate) scale, rather than in a sharpness (univariate) scale, the results were rescaled to convert quality steps to sharpness steps, by employing a validation test. In calibrated perceived sharpness scale, the average difference in perceived sharpness was approximately 11.86 JNDs with 75% change in the displayed image size. The red dots in both Figures 5-14 was estimated by linear extrapolation, and was based on the rather

assumption that the response of the visual mechanisms remains the same when images are displayed in larger sizes.

Chapter 6

Psychophysical investigation 3: Evaluation of changes in perceived contrast with changes in displayed image size

This chapter is concerned with a quantification of the degree of change in the perceived image contrast with respect to changes in displayed image size. This was achieved by collecting data from psychophysical investigations that used techniques to match the perceived contrast of displayed images of five different sizes using the method of adjustment in a dark environment. The chapter also details a method employed to create a series of S-shaped filters, which were implemented in the spatial domain and were designed to provide 25 equal intervals in global perceived image contrast. In addition, the validation of results obtained from the contrast matching experiments, the

evaluation of the step intervals and the calibration of the gamma scale to a contrast JND scale were described.

6.1 Introduction

The contrast of reproduced scenes depends on the tone reproduction of the imaging systems employed. Fairchild (Fairchild, 1995) described objective contrast as “the rate of change of the relative luminance of image elements of a reproduction as a function of the relative luminance of the same image elements of the original image”.

Perceived contrast is, however, a visual phenomenon. Even if the visual contrast is dependent upon the objective contrast and affected by the absolute luminance levels of the image being viewed (Giorgianni and Madden, 2008, p.26), it is greatly influenced by the background (and the surround) (Fairchild, 2005, p.111-127).

Braun and his colleague have remapped lightness using sigmoid functions to enhance image contrast based on the phenomenon of simultaneous lightness contrast (Braun and Fairchild, 1999). Image appearance is known to be affected by the background (and the surround) (Hunt, 1952). Thus, it is possible to make the highlight image area in an image appear lighter by making the shadow areas darker, which results in an increase in the perceived image contrast. This technique is based on the knowledge that the human visual system does not work on an absolute basis but instead it works on a relative basis (Giorgianni and Madden, 2008, p.26). In other words, the human visual system is more sensitive to contrast rather than absolute luminance.

In LCD systems, tone reproduction is defined as the functional relationship between the input pixel values and the output luminance, and contrast can be expressed by gamma, γ . When the relationship is plotted in linear units and described by a power

function, the exponent represents gamma (c.f. Section 2.2.2.1). Bilissi *et al.* (Bilissi *et al.*, 2008) have conducted various psychophysical experiments to evaluate acceptable and just perceptible gamma differences using cathode ray tube (CRT) displays under both controlled and uncontrolled environments. The just perceptible differences in gamma were 0.12 and 0.10 under controlled and uncontrolled environments, respectively.

The purpose of the creation of the filters was to produce test images with different contrast and thus enabling us to quantify the changes in perceived image contrast with respect to changes in displayed image size. In this task, it was essential to take into account the perceptual gamma differences whilst keeping the mean image luminance unaltered.

6.2 Preparation of test stimuli

6.2.1 Creation of a series of contrast filters with n-JND interval

In order to create a set of filters to increase the image contrast and their corresponding inverse functions, the S-shaped filter functions were manually created. For this work, a set of twenty four filter functions were created using the following steps. The step intervals were calculated by adjusting the gamma of the input to output transfer curve.

1. Pixel values (PV) ranging between 0 and 128 (half way the pixel values range) were selected and normalised (divided by 128).
2. Corresponding output PVs were calculated using a power function with exponent (gamma, γ), ranging between $\gamma=1.6$ and $\gamma=1/1.6$ with intervals of 0.05

gammas (approximately half a perceptible gamma difference) (Bilissi *et al.*, 2008).

3. Normalised original and corresponding PVs were converted back to their original range (0 to 128).
4. Corresponding output PVs were then mirrored at PV of 128 for the calculation of PVs between 128 and 255.
5. 6th order polynomials were fitted to the calculated output pixel values using Oakdale Engineering Datafit v9.0 (Oakdale Engineering, 2007).
6. Actual gammas of each function were obtained for the mid-tones (PV between 96 and 160).

Filter functions for the gamma adjustment are illustrated in Figures 6-1 and 6-3.

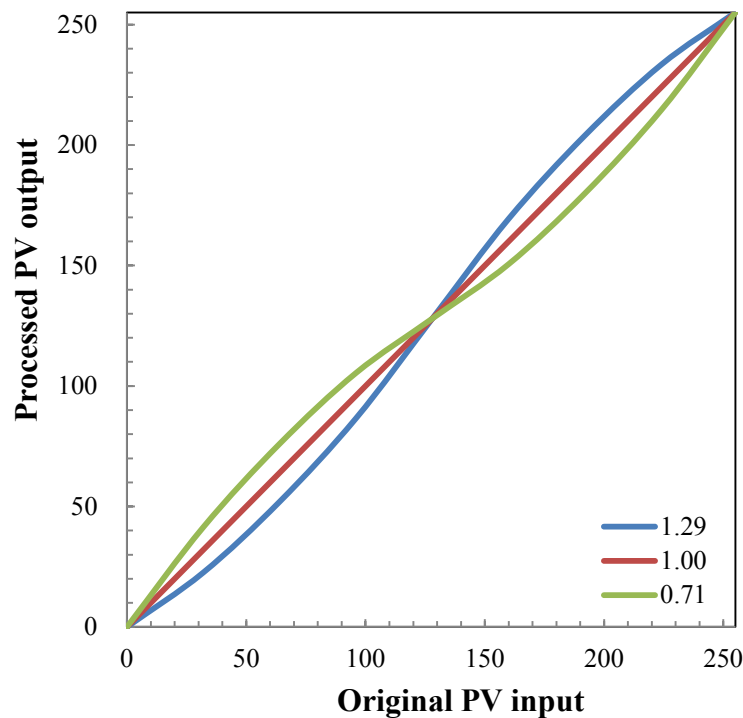


Figure 6-1. Sample S-shaped filter functions, calculated by gamma adjustment by power transformation.

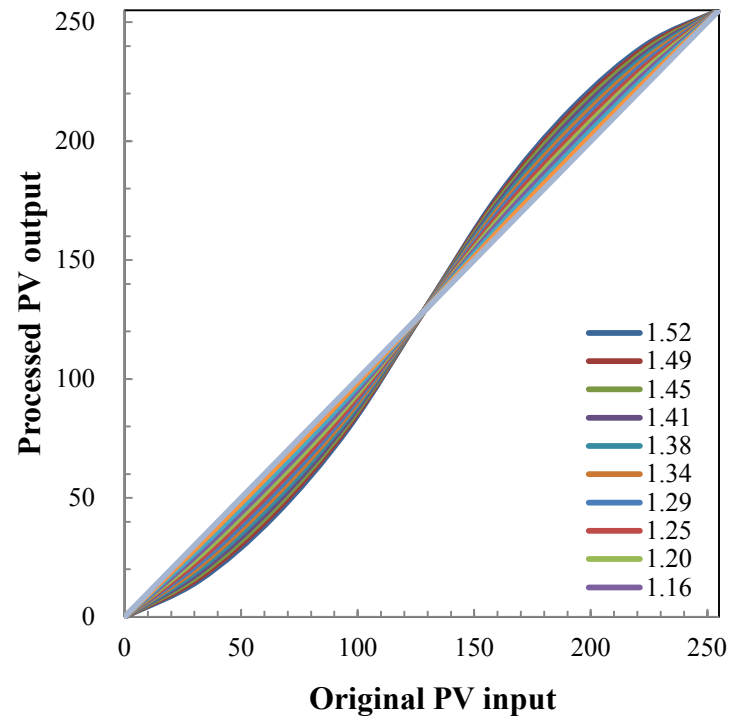


Figure 6-2. A series of gamma increasing filter functions.

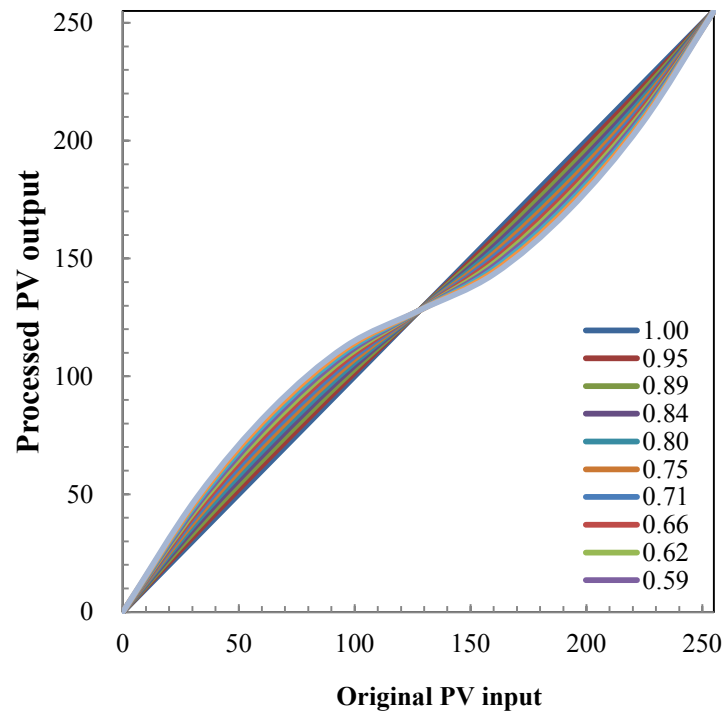


Figure 6-3. A series of gamma decreasing filter functions.

6.2.2 Spatial domain filtering

The filtering operation was carried out using MATLAB. The filter functions were applied directly on each pixel of the sixty-four original version images on the R, G, and B channels. A total of 25 ruler images, each possessing different perceived contrast level with equal gamma difference (original, 12 contrast decreased versions and 12 contrast increased versions), were generated in spatial domain. Filtered images were then resized to five different versions by bi-cubic interpolation. The changes in mean luminance of the images were not evident. The dimensions of the resized test images were identical to those described in Chapter 5.

Sample image and its filter versions were present with image histograms in Figure 6-4.

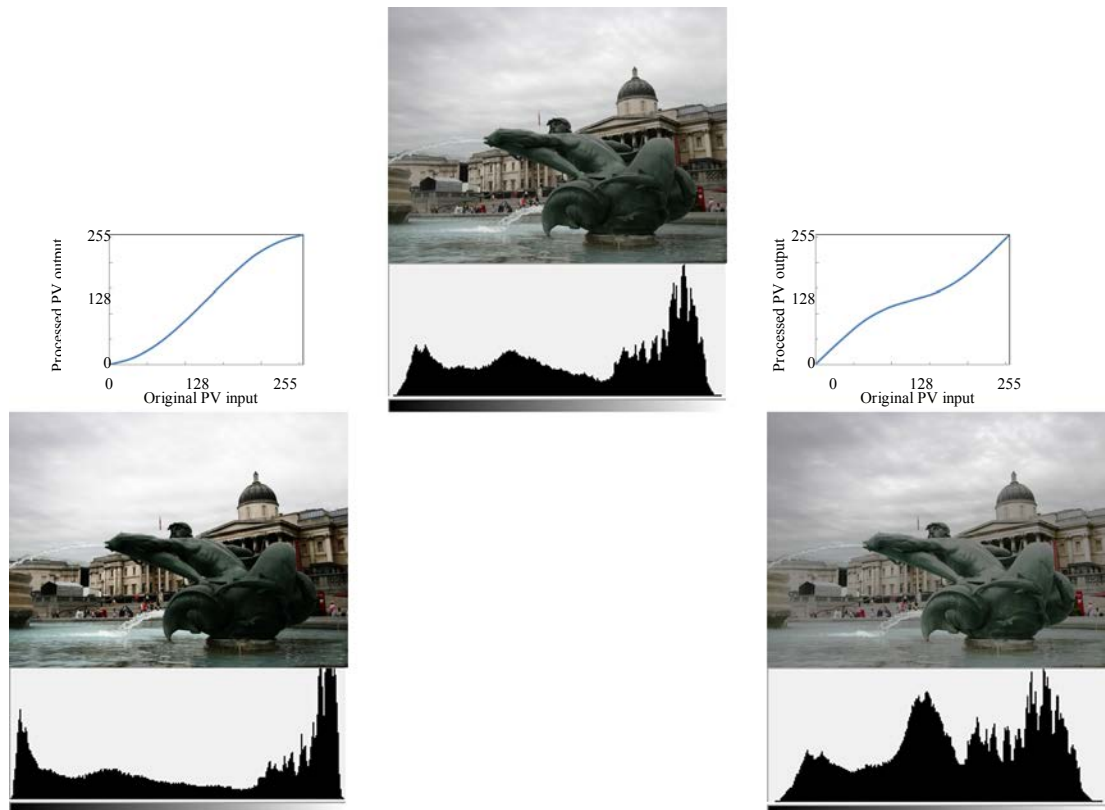


Figure 6-4. Sample S-shaped filters and the contrast manipulated images. Original image (top), contrast increased version at $\gamma=1.52$ (bottom left) and contrast decreased version at $\gamma=0.48$ (bottom right).

6.2.3 Contrast measurement of the ruler images

In order to confirm the contrast changes in ruler images objectively, the root mean square (RMS) contrast, which is one of the most commonly employed metric for the purpose, was measured (Peli, 1990). RMS has been shown to correlate successfully with human contrast detection not only for the laboratory stimuli but also for natural images (Bex and Makous, 2002, Frazor and Geisler, 2006). RMS contrast is defined by the root mean square deviation of the pixel luminance from the mean pixel luminance of the image, divided by the image dimension (Pavel *et al.*, 1987). RMS contrast, C_{RMS} , of a two dimensional image are defined in Equation 6.1, adapted from Peli (Peli, 1990).

$$C_{RMS} = \sqrt{\left[\frac{1}{R \times C} \sum_{x=1}^{C-1} \sum_{y=1}^{R-1} (I_{xy} - \bar{I})^2 \right]} \quad (6.1)$$

where R and C are the number of rows and columns in the image, I_{xy} is the normalised luminance of x^{th} y^{th} pixel, \bar{I} is the mean normalised luminance of the image.

C_{RMS} of all sixty-four test images and that of their ruler versions were measured in display luminance space. Each original scene possessed a different C_{RMS} value and the degrees of change in C_{RMS} differed on ruler versions of each scene. However, changes in C_{RMS} on filtered images showed a linear trend. C_{RMS} values of four selected images of the large version are plotted in Figure 6-5 for illustration purposes. The selected scenes include those possessing the highest C_{RMS} and the lowest C_{RMS} and two scenes possessing average C_{RMS} .

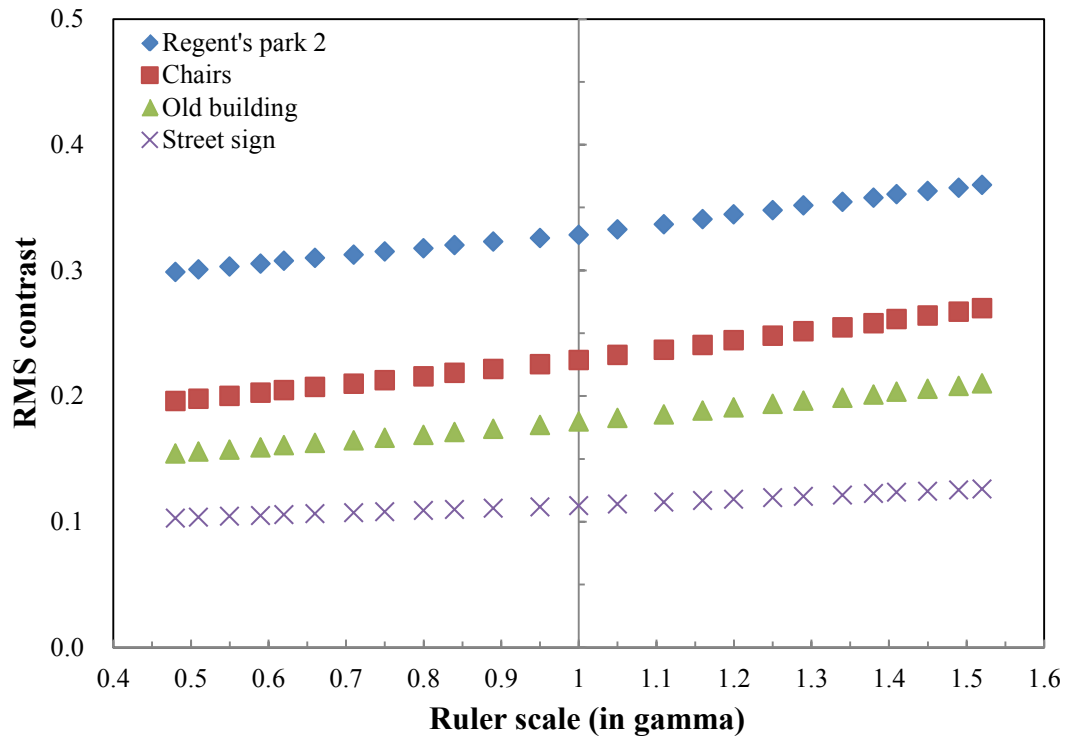


Figure 6-5. C_{RMS} of four selected scenes at a different ruler scale.

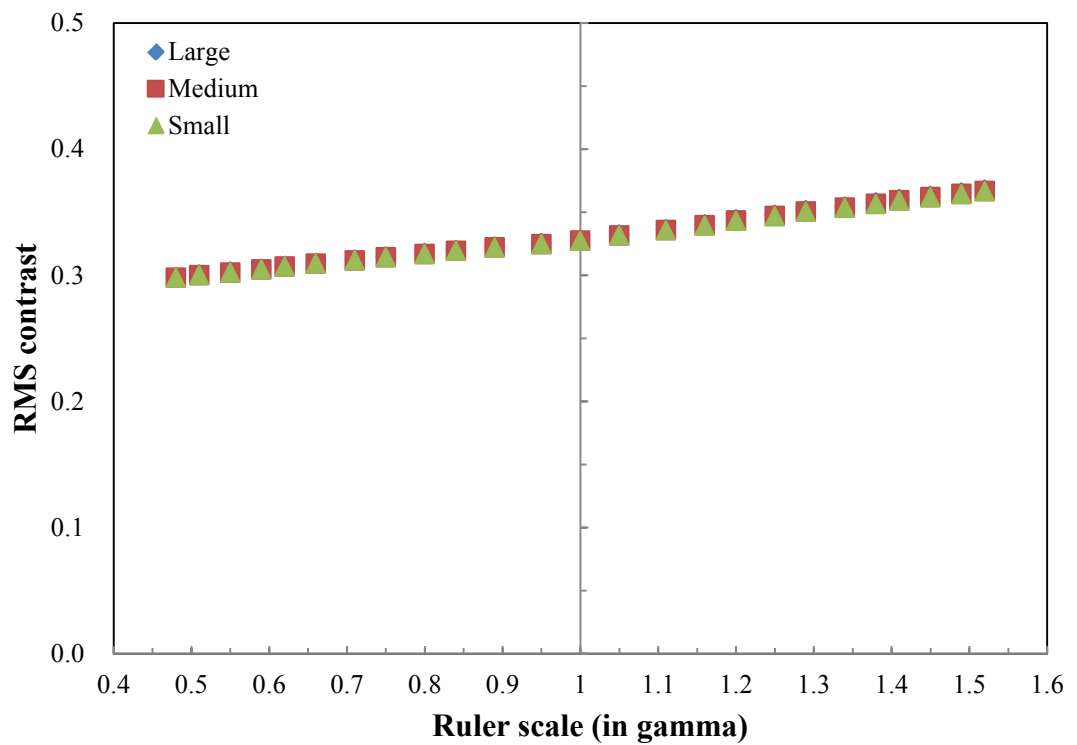


Figure 6-6. C_{RMS} of 'Regent's Park 2' at a different ruler scale in 3 different image sizes.

In addition, the effect of bi-cubic interpolation on the measured image contrast was investigated. C_{RMS} of all test images at five different sizes were measured. However, the effect of bi-cubic interpolation on C_{RMS} was not evident. C_{RMS} of the filtered ‘Regent’s Park 2’ scene at various image sizes are shown in Figure 6-6.

6.3 Psychophysical investigation

Visual contrast matching tests, using a slider controlled by the computer mouse, were also conducted in a totally dark environment, as described in Section 5.2.3. The same display, settings, calibration, and user interface were used as for the sharpness matching experiment (c.f. Section 5.2).

Observers were seated on a comfortable seat with a chin rest to hold the observation distance at 60cm from the display. Observers were requested to move their eyes from side to side only. During the tests, a randomly selected test image was displayed simultaneously at two different sizes. The test images were displayed in random display sides, one on the left side and the other on the right side of the display. Observers were asked to match the contrast of the smaller ‘test’ images to that of the larger ‘standard’ images using a slider. The slider was programmed to simulate to the user an enhancement of contrast of the images in response to changes in the slider position by replacing the test image with the appropriate ruler image, according to the selected slider position.

Experiments consisting of four contrast matching sessions were carried out: small size to large size, medium-small size to large size, medium size to large size, and large-medium size to large size. A total of twenty observers, 5 females and 15 males, participated in the experiment using all 64 scenes. Their age ranged between 20 and 40

years old and all of the observers had imaging and design backgrounds. Each observation took less than one hour per session; one session per day was conducted to avoid fatigue.

6.4 Results and discussion

The mean, μ , and standard error of the mean (SEM) were calculated for each scene and size pairs.

6.4.1 Results from the psychophysical tests

Observations from matching the contrast of the small version image to that of the large version image, resulted in an average change in tone reproduction of 0.087 gamma (or 2.0 steps in the contrast scale), with an average standard error of mean (SEM) of 0.030. The range of change for all scenes was from -0.04 to 0.19 . The changes for each scene, along with standard error, are plotted in Figure 6-7.

From the experiment of the medium-small against large pairs of images, the average change was 0.050 gamma with an average SEM of 0.027. The range of change was from -0.08 to 0.14 . From the medium version against the large version image, the average change was 0.043 gamma with an average SEM of 0.022. The range of change was from -0.02 to 0.13 . And the large-medium version against the large version matching experiment showed that the average change was 0.036 gamma with an average SEM of 0.023. The range of change was from -0.054 to 0.096 . The results are plotted in Figures 6-8 to 6-10.

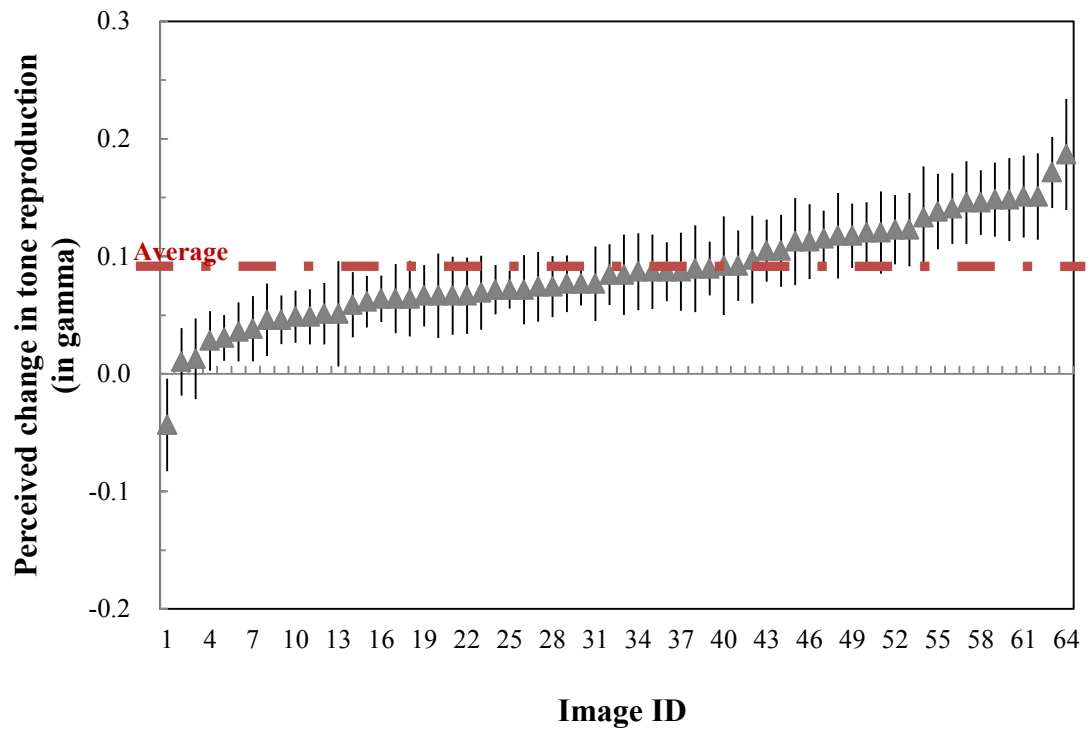


Figure 6-7. Average perceived change in tone reproduction from the small vs. large experiment for each scene with SEM.

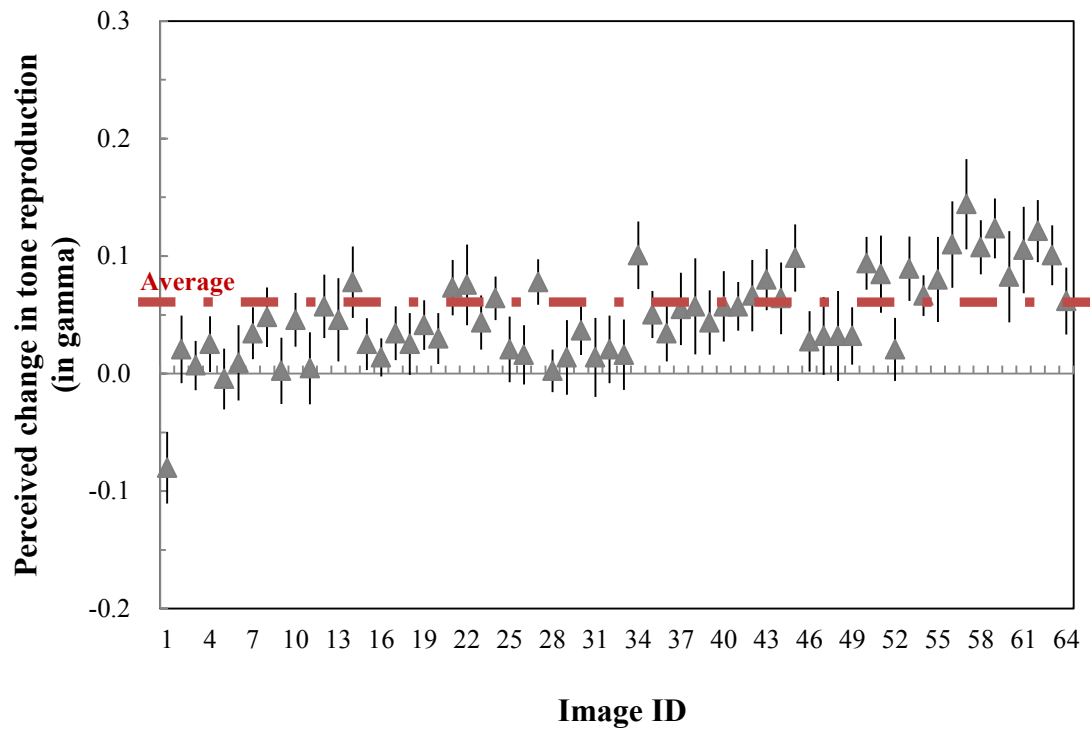


Figure 6-8. Average perceived change in tone reproduction from the medium-small vs. large experiment for each scene with SEM.

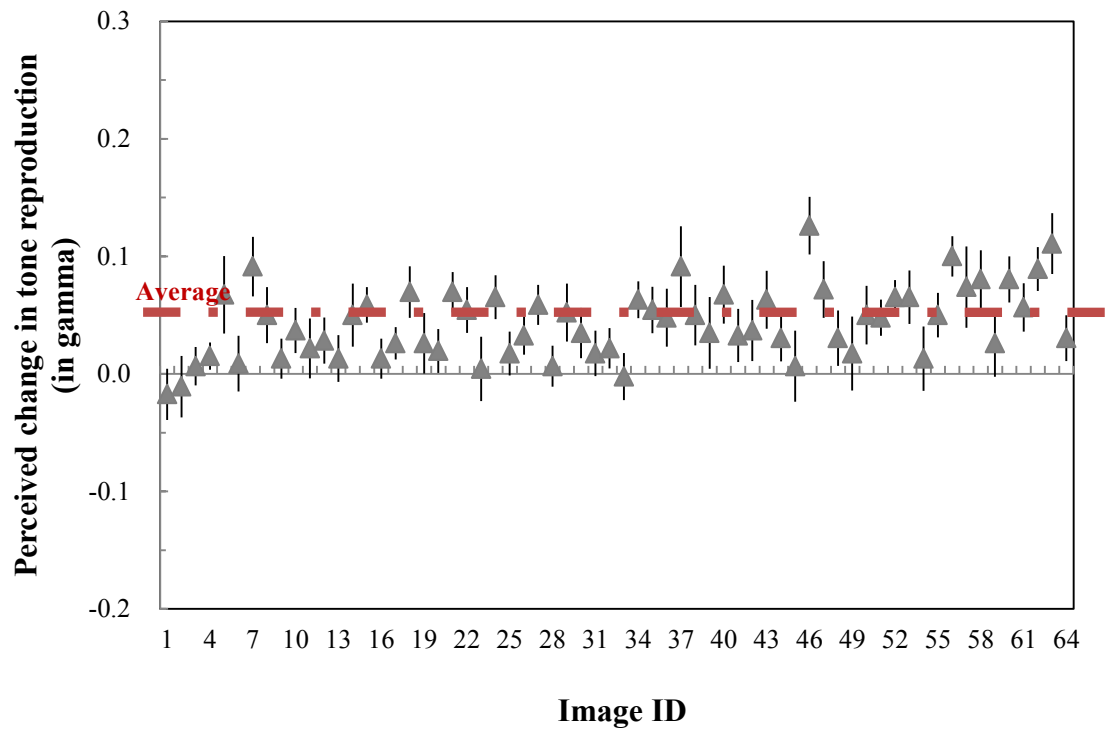


Figure 6-9. Average perceived change tone reproduction from the medium vs. large experiment for each scene with SEM.

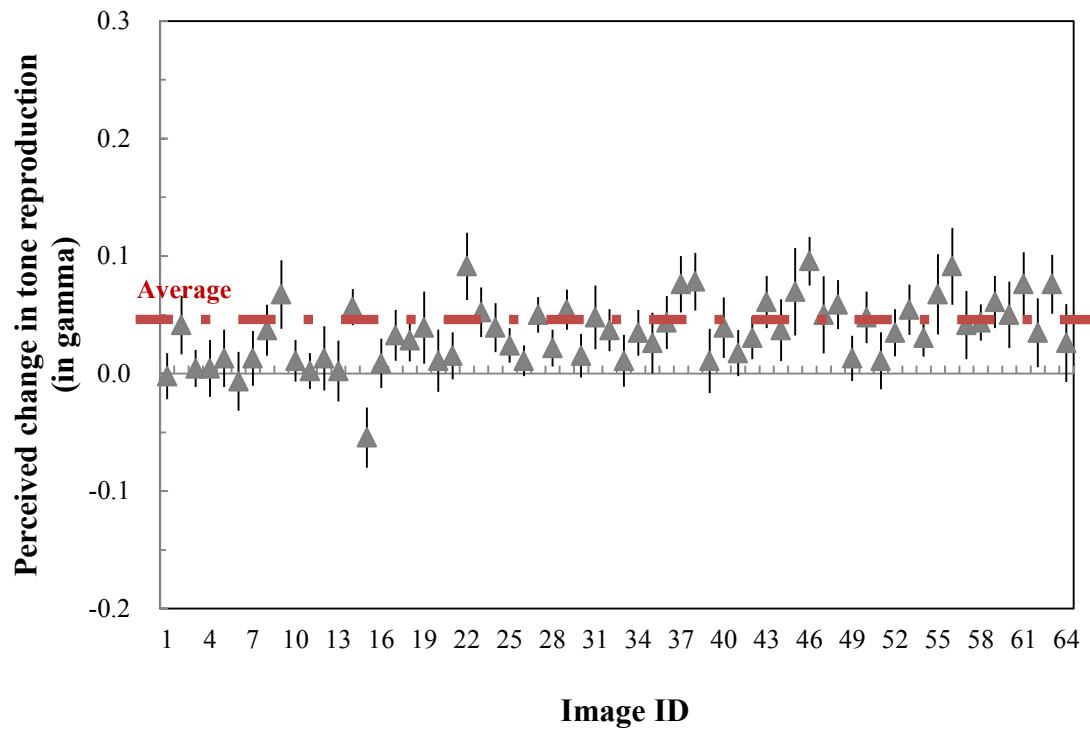


Figure 6-10. Average perceived change in tone reproduction from the large-medium vs. large experiment for each scene with SEM.

In addition to the above figures, the average changes in perceived tone reproduction in gamma from all four experiments were plotted as a function of displayed image size in Figure 6-11. The figure clearly shows that the perceived contrast was proportionally affected by the changes in displayed image size. Smaller version images were perceived a higher contrast than that of the larger version and their relationship was very close to an inverse linear relationship as seen in the previous chapter. Therefore, mirrored data at zero point have also been estimated by extrapolation and plotted as linear function to predict change in perceived contrast when images may be displayed at larger scales. This assumes that the relationship remains linear. The linear trend line showed the relationship as; $y = -0.001x + 0.0005$.

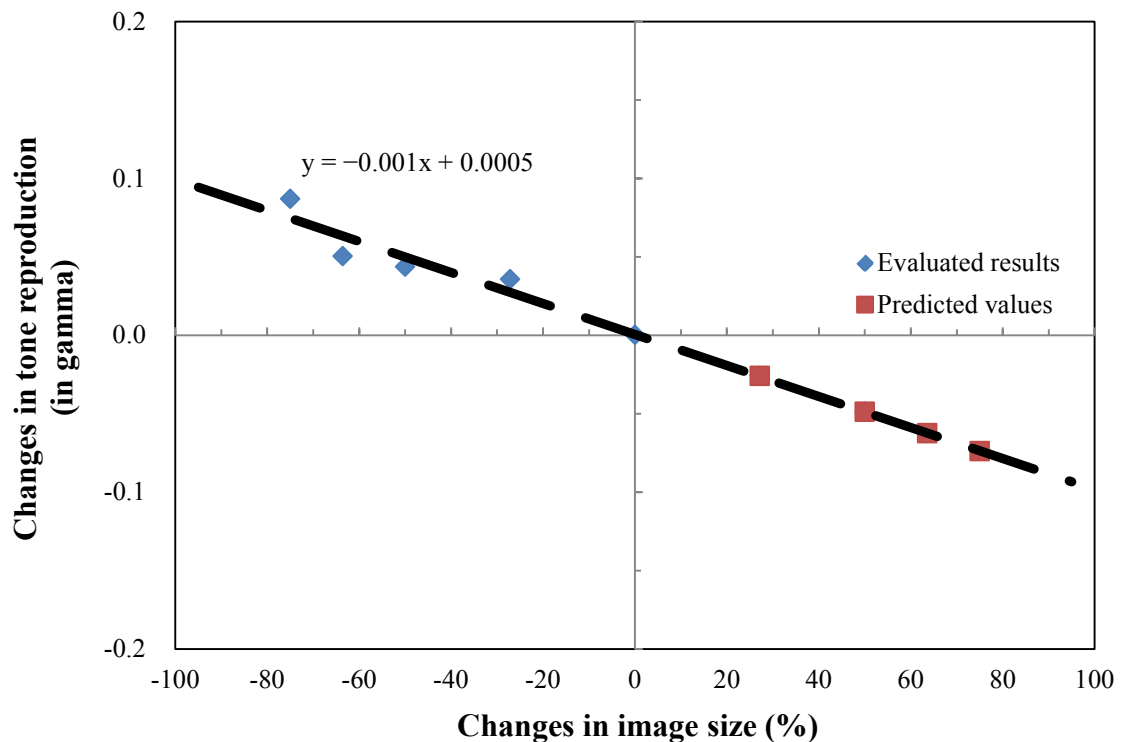


Figure 6-11. Perceived changes in tone reproduction with respect to the changes in displayed image size (blue) and predicted changes (red) in non-calibrated relative image quality gamma scale.

6.4.2 Validation of the results

Pair rating experiments, to validate the results obtained from the contrast matching, were conducted and analysed in conjunction with the validation of the results obtained from the sharpness matching. Therefore in addition to the test images prepared in section 5.3.2), the contrast modified smaller version images were prepared.

Results from the validation experiments confirmed that most of the contrast matched pairs, average rating of 4.90, appeared to be better matching compared with that of the original pairs, 4.62, as shown in Figure 6-12.

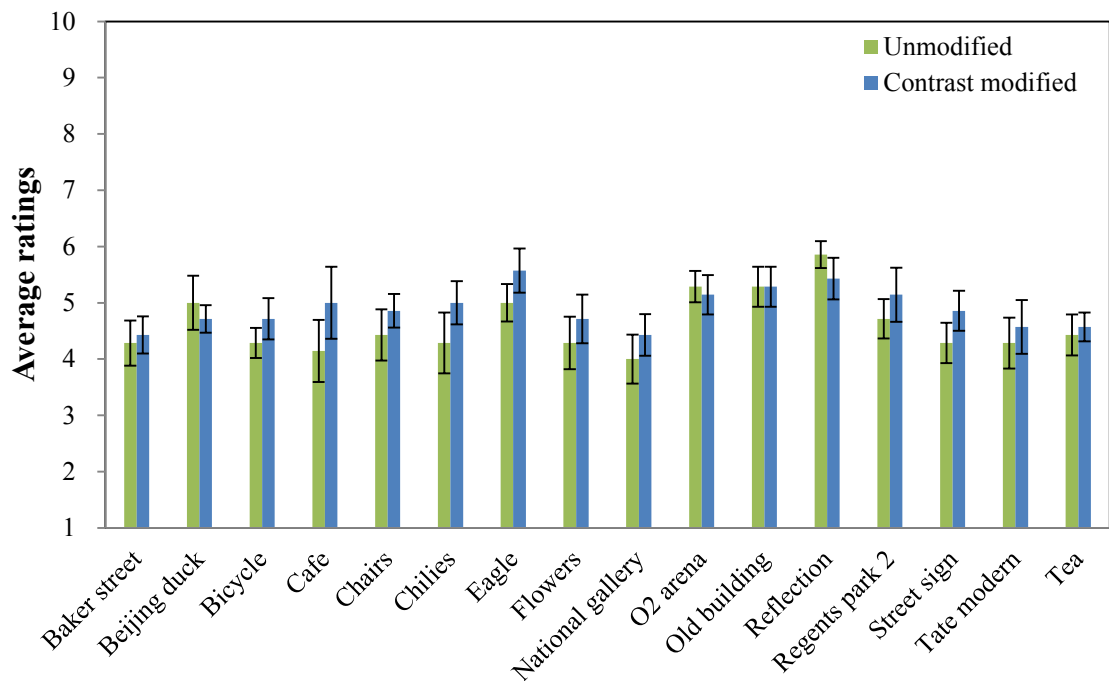


Figure 6-12. Average rating of the unmodified pairs and the contrast modified pairs.

6.4.3 Evaluation of step interval and calibration of changes in contrast JND scales

In order to validate the results acquired by the contrast matching investigations, a series of pair rating experiments was conducted under the same experimental environment, as

described in Section 4.2. A series of paired comparison experiments to evaluate the contrast step intervals were conducted using all sixty four scenes. For the contrast step interval evaluation, only the central region of the scale (original ± 6 steps) was used, as most of the appearance changes were found within the range. Experiments were carried out by the expert observers who have also participated in sharpness step evaluation. Each observation took less than 10 minutes per session and a maximum of 10 sessions per day was conducted to avoid fatigue.

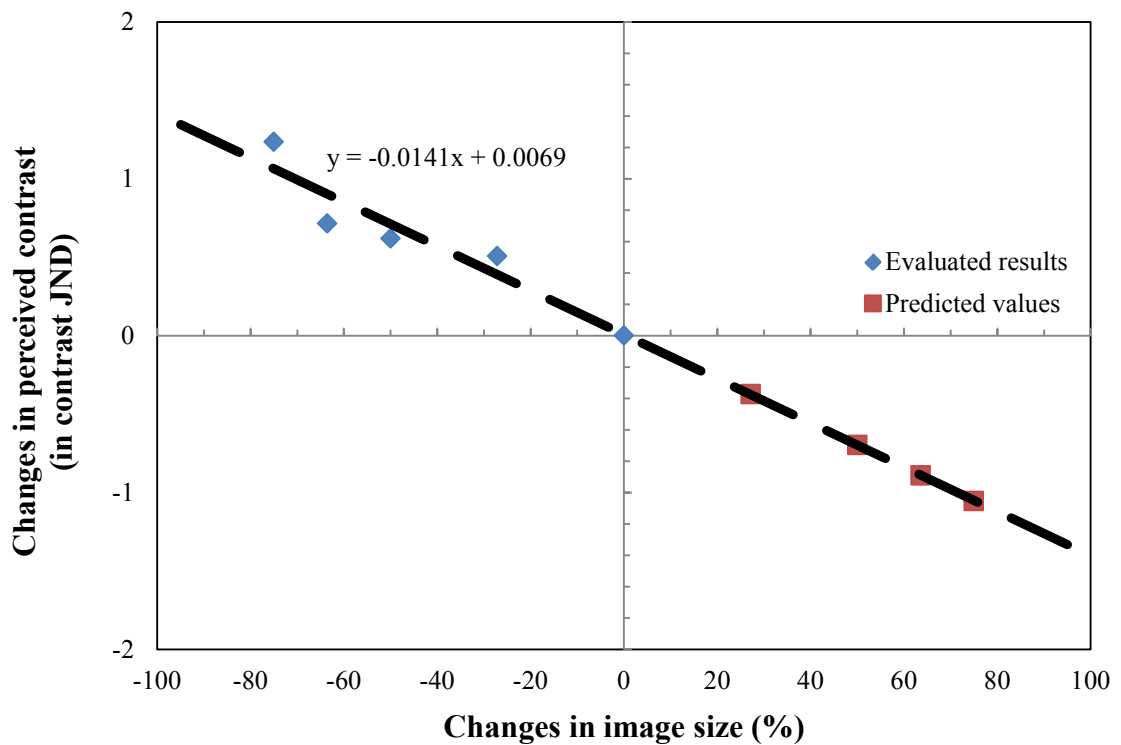


Figure 6-13. Changes in perceived contrast with respect to the changes in displayed image size (blue) and predicted changes (red) in contrast JND scale.

The outcome was used to calibrate the results plotted in Figure 5-14 into perceptually meaningful scales for the sharpness attribute. From the experiments for the sharpness step validation, an average of 0.070 gammas (or 1.614 steps) in the ruler scale were found to be 1 JND in perceived contrast. The results obtained in Section 6.4.1

were then calibrated and plotted in Figure 6-13. The change in perceived contrast was approximately 1.24 JNDs with 75% change in the displayed image size. The linear trend line showed the relationship as; $y = -0.014x + 0.006$.

6.5 Summary

Since contrast was identified in Chapter 4, as the second most affected image attribute when displayed image size changes, a series of psychophysical experiments were carried out to evaluate the changes in perceived contrast when the images are viewed at different displayed sizes on an LCD device. A total of sixty-four natural scenes, which were used for the experimental work described in Chapters 4 and 5, were also used in this investigation. For each original scene, a set of 25 images of varying image contrast with equal gamma interval were created, using 'S' shaped 6th order polynomials. The processed images were resized to generate five different sizes: large, large-medium, medium, medium-small and small using bi-cubic interpolation. As for the sharpness matching, the observers started to match smaller version to the large reference. Results from all four experiments showed that most for majority of the test scenes, the smaller version images were perceived as slightly more contrasty compared with the reference images. A minority of test images did not show the same trend. Examples include 'British museum' scene, which contained large amount of the dark reflection (Normalised PV of less than 0.5). Although overall gamma of the processed images was decreased when contrast decreasing filters were applied, such compressed shadow details became more evident.

Overall results from the psychophysical experiments indicated that the perceived contrast was affected by changes in displayed image size; however, it was

much smaller compared with the changes in perceived sharpness. Average difference was approximately 0.087 gamma between large version images and small ones.

Also, the changes were quantified and presented in a “gamma” scale rather than in a perceived contrast scale from the contrast matching experiments. Therefore, paired comparison experiments to evaluate the step intervals of the contrast were carried out. The results acquired from the contrast matching tests were rescaled to contrast scale by step validation test. In calibrated scale, average difference was approximately 1.24 JNDs in perceived contrast with 75% change in the displayed image size. This is a rather insignificant visual difference when compared to the sharpness difference. We may thus conclude that the visual contrast difference produced by changing displayed image sizes probably do not affect significantly the overall quality of the images.

Chapter 7

Discussion

This chapter provides a summary of the characterisation of imaging devices used for image capture and display, and discusses their effects and limitations with respect to the psychophysical experiments carried out in this research project. Detailed discussions on the results from the psychophysical investigations described in Chapters 4, 5, and 6 are also included.

7.1 Capturing devices

In this research, two digital cameras, exhibiting difference overall image qualities, were used for the image capture of natural scenes. The Canon EOS 30D digital SLR was

equipped with an EF-S 10-22mm (35mm equivalent focal length of 16-35mm) lens. The Apple iPhone mobile phone camera (1st generation) was equipped with a fixed lens (35mm equivalent focal length of 35mm). Although Canon 30D allowed full access to camera functions, this was not allowed on the Apple iPhone. This fact restricted us of implementing the most accurate characterisation methods (such as spectral characterisation). Tone reproduction and colorimetric characteristics of the capturing devices were carried out for the sRGB setting, using target-based methods for both cameras, for consistency. In addition, the Spatial Frequency Responses (SFRs) of cameras were measured under identical conditions. Under the experimental set up for the camera characterisation, both systems exhibited difference characteristics and overall quality, with the Canon 30D exhibiting the better quality. Colour reproduction of both devices was not accurate with colour differences of over $10 \Delta E_{ab}^*$ from both devices. Greater differences were observed from the ‘reddish’ patches reproduced by both devices, as seen in Figure 3-3. The maximum colour differences were as high as $33.14 \Delta E_{ab}^*$ and $47.23 \Delta E_{ab}^*$ on reddish patches by the Canon 30D and the Apple iPhone, respectively. The main cause of such inaccurate colour reproduction was the small difference in the white point colour temperature of the light source (2700K) and the colour balance of the cameras (3200K). Despite these large colour differences, the image quality was not visually affected by the slight failure of colour balance. Since the purpose of most commercial digital camera is to prioritise pleasing reproduction over colorimetric reproduction, also since the research was not focused on colour appearance, the colorimetric inaccuracies in the captured images did not affect the image quality of the selected test stimuli, or the design of the psychophysical experiments.

The sharpness characteristics of both cameras were assessed by SFR evaluation. SFRs were measured using the slanted edge method (ISO 12233). Enhanced edge sharpening was evident on the SFRs obtained for the Apple iPhone. Also, SFRs varied considerably between measurements and channels, even within the same test environment. On the other hand, the SFRs obtained by the Canon EOS 30D camera indicated repeatable results, whilst the variations between channels were fairly small. Consistent SFR measurement was essential for this research, since the camera SFR was used to predict the sharpness of the original test stimuli, and further employed in the creation of a series of test stimuli with different sharpness levels.

7.2 Display devices

In this research, two LCD monitors were used to display test stimuli during the psychophysical investigations. The purpose of employing LCD devices in psychophysical investigations is to display test stimuli with good positional uniformity, rather than accurate colorimetric reproduction. Originally, the EIZO ColorEdge CG210 display was used in the investigation. Although the CG210 exhibited a good black level and accurate reproduction of primarily colour at full strength, positional non-uniformity nature of the display was considerably large. Since the positional independence was the main concern for the experimental works conducted in this research, the CG210 was not good enough for further investigations. Therefore, a new EIZO ColorEdge CG245W display was employed for the further experiments.

The CG245W exhibited better characteristics compared with the CG 210 display in all aspects, including positional uniformity. Average colour reproduction error, ΔE_{ab}^* , was 1.53 across the screen with maximum error of 3.89 (at the edges). As

the psychophysical investigation was carried out using standard reference images with a horizontal visual angle of approximately 20 degrees at the set observation distance, positional uniformity characteristics were further evaluated at the observation position. The colour reproduction error, ΔE_{ab}^* , was slightly bigger with an average of 2.53. Maximum error of $\Delta E_{ab}^*=5.24$ was found at the edges. Although the errors were rather bigger, perceptible and acceptable colour differences are subjective and their significance depends on the application. Theoretically, 1 ΔE_{ab}^* is approximately 1 JND. However, display system with a ΔE_{ab}^* of smaller than 6 is commonly accepted for displayed images (Abrardo *et al.*, 1996).

7.3 Identification of image attributes

The purpose of the experimental works described in Chapter 4 was to identify the image attributes that were most affected visually by changes in displayed image size. Other workers have previously conducted research in attempts to identify the image attributes that are affected by changes in image size, or visual angle. Research work by Choi and her colleagues (Choi *et al.*, 2007b) confirmed that perceived colourfulness was affected by changes in colour patch size as well as changes in viewing conditions such as surround and relative luminance. Nezamabadi and his colleagues confirmed that perceived contrast (Nezamabadi *et al.*, 2007), lightness, and chroma (Nezamabadi and Berns, 2006) were affected by the changes in visual angle and perceived image size. However, spatial effects and appearance of digital image artefacts were not considered in depth.

Therefore, in this research, a novel psychophysical experiment was designed to investigate spatial effects such as sharpness and noisiness, along with other colour

attributes. The experiment was carried out using natural scenes with different scene content, taken under various illumination conditions. Investigated attributes in the forced choice experiments included contrast, brightness, sharpness, and noise for both achromatic and chromatic versions of the stimuli. Hue and colourfulness were also investigated for chromatic version stimuli. Various approaches were made to analyse the data obtained by the experiments to identify the effects of scene characteristics, such as average scene luminance, colourfulness, busyness, sharpness, and noisiness. The results differed slightly when analysed according to the scene characteristics listed above. However, the first two, most affected attributes, were found to be the first two for the majority of the test scenes.

Results from the rank order experiments using achromatic stimuli showed that the most affected image attributes with respect to change in displayed image size were sharpness followed by contrast. Experiments using chromatic versions confirmed these results.

Contrast is considered as the most important aspect of image quality (Triantaphillidou, 2011a, p.346, Hunt, 1998), whilst sharpness is directly related to the micro-image (edge) contrast (c.f. Section 2.2.2.4) as well as to the image's angular subtense. The fact that these two attributes were found to be the most affected ones when changing displayed image size is thus not a surprising finding.

7.4 Sharpness matching

The purpose of the experimental works described in Chapter 5 was to quantify the degree of change in perceived image sharpness with respect to changes in displayed image size. A novel method was designed to create a range of images with varying

sharpness levels, by adapting the softcopy ruler method (*Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method*. 2005), as described in Chapter 5. A series of filters were created for the purpose, having equal intervals in image quality, by taking into account the SFR of the imaging system. This method assumed that images acquired by the same capturing device and using identical lens settings (i.e. aperture and focal length) would have the same SFR. Effects of scene content and illumination conditions during image capture on image sharpness were not taken into account.

A series of visual sharpness matching experiments were carried out using the method of adjustment. Results from the sharpness matching experiment showed that all test images were perceived sharper when image size was decreased. In other words, perceived sharpness may decrease when image size increases (c.f. Section 5.3.1). The results suggest that when images are viewed in the small camera displays just after capture, they are likely to appear much sharper, in most cases, than when they are viewed later at a 1:1 magnification on a computer display. This is a common experience of camera users. The effect is particularly important when the lack of sharpness is due to camera, or object movement introduced during capture. Images that included either moving objects or camera shake were less affected, and images that included texts or repeated objects were most affected (Park *et al.*, 2012).

A psychophysical experiment to validate the results obtained from the sharpness matching experiments was also conducted (c.f. Section 5.3.2). On some images, unmodified versions were perceived to be closer matching. However, majority of the sharpness modified small versions were perceived to be closer matching to the large original than the unmodified small versions for the majority of the images, as

shown in Figure 5-15. Further, in some cases the error bars however overlap, making it unclear whether the modified sharpness of the small images was clearly better than the unmodified original. Further work is needed to identify the original sharpness characteristics of the scenes with close results, also to determine whether the average sharpness JND that was applied as a correction in this validation stage was rather simplistic solution.

7.5 Contrast matching

The purpose of the experimental work described in Chapter 6 was to quantify the degree of change in perceived image contrast with respect to changes in displayed image size. A method of creating a series of S-shaped spatial domain filters for contrast manipulation, with equal gamma intervals, was described in Chapter 6. The step intervals were selected by adapting acceptable and just perceptible gamma differences evaluated using CRT displays by Bilissi *et al.* (Bilissi *et al.*, 2008), for a small image size of 75(H)×112(V) mm that corresponded to an angle of subtense of approximately 15 degrees.

A set of four visual sharpness matching experiments were carried out using the method of adjustment. Results from the contrast matching experiment showed that perceived contrast was increased when image size was decreased, what was also observed in the sharpness matching experiment. In other words, perceived contrast may also decrease when image size increases (c.f. Section 6.4.1).

A psychophysical experiment to validate the results obtained from the contrast matching experiments was also conducted (c.f. Section 6.4.2). As shown in Figure 6-12, the majority of the large original-contrast modified small version image pairs rated

superior compared with the large original-small unmodified image pairs. On some images, unmodified version images were perceived to be closer matching to the large original. The error bars indicate that the difference between the modified and the unmodified contrast in small images is not as important as for in the case of sharpness.

Chapter 8

Conclusions and recommendations for further work

This chapter contains conclusions drawn from this research project along with recommendations for further work.

8.1 Conclusions

The following conclusions were drawn from the research work conducted in this thesis:

- Image attributes affected visually by changes in image size in softcopy reproduction were identified using natural scenes. Two camera systems were employed to capture the same scenes to investigate the effect of original image quality on image appearance. Six image attributes were investigated by ranking experiments. Results varied slightly with scene content and original image quality characteristics. However, sharpness and contrast were identified

as two of the most affected attributes for the large majority of scenes, followed by contrast and brightness.

- A series of filters were successfully created to create a series of image with equal intervals in image quality. This was done by taking into account the Spatial Frequency Response (SFR) of the imaging system and by adapting ISO 20462-3.
- The effect of bi-cubic interpolation on image quality was investigated also via SFR measurements. SFRs of the interpolated versions of a number of images were found to be lower than the SFR of the larger reference version.
- Matching experiments with images displayed in different sizes showed that perceived sharpness increased when image size was decreased. Test images containing either moving objects or camera shake were less affected and images that included texts or repeated objects were more affected by image size changes. Changes in image appearance between the smaller version images and the larger versions had an average of approximately 12 sharpness JNDs.
- The effect of bi-cubic interpolation on image contrast was investigated by measuring the root mean square contrast, C_{RMS} , on all interpolated version images and was compared to the C_{RMS} of the original test image. Each test image possessed a different C_{RMS} . However, the effect of interpolation on contrast was minimal for all images. Although C_{RMS} was not affected by change in image size, perceived contrast increased when image size decreased on the majority of test images. Changes between smaller version images and larger reference were approximately 1 contrast JND.

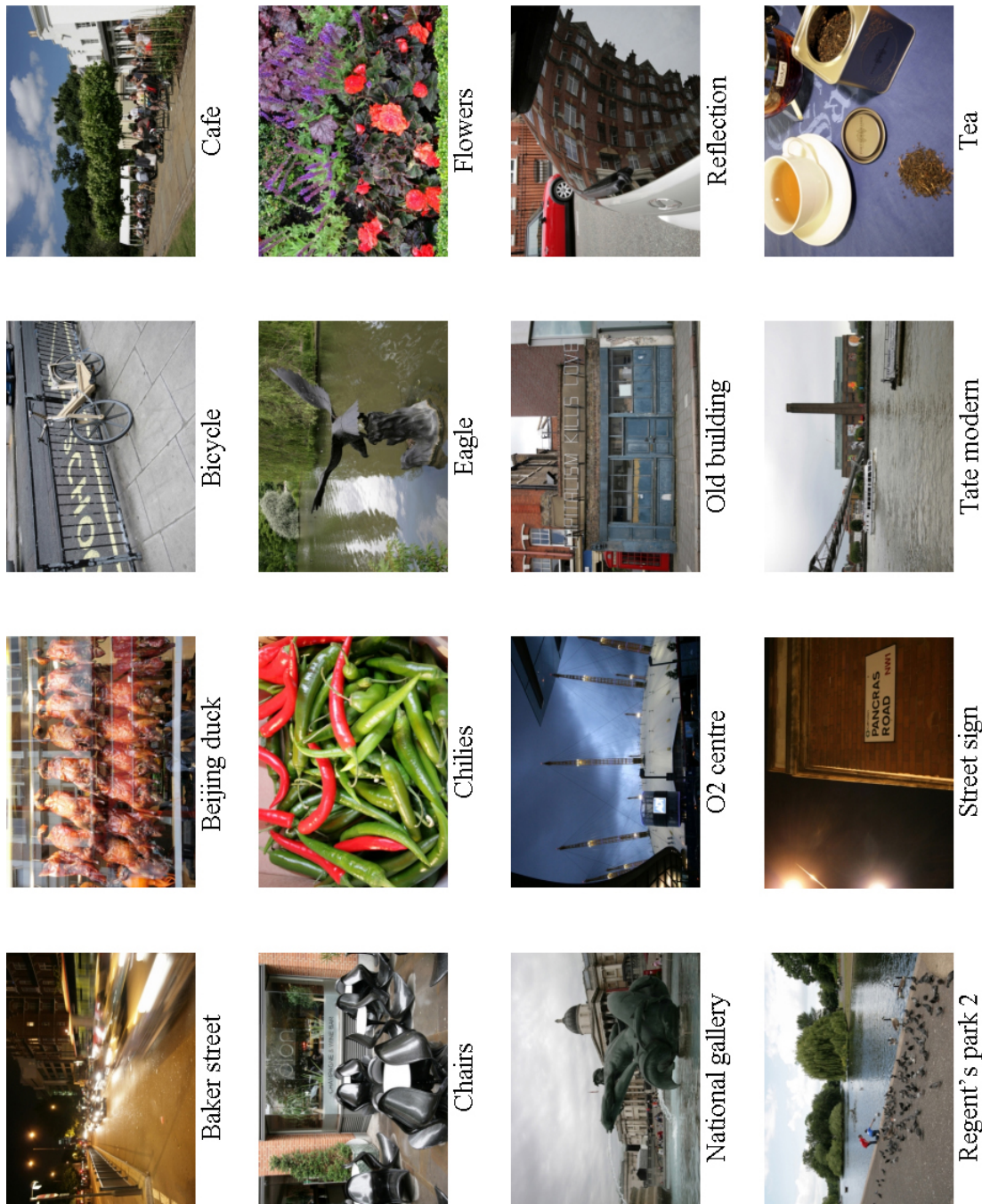
8.2 Recommendations for further work

Some recommendations for further work were as follows:

- Perceived sharpness and contrast in complex pictorial images were investigated in this work to identify how these attributes were affected by change in displayed image size. However, the appearance of colour attributes (such as colourfulness and hue), noise, as well as various image artefacts (such as blocking, banding and aliasing) have not been researched here. Further work could include such investigations, relating image size and appearance of these attributes and artefacts with changes in image size, which to the author's knowledge have not been studied in depth up to date.
- Investigations were carried out in a totally dark environment. Effects of the surrounding viewing conditions on image appearance, with respect to change in displayed image size, can be investigated further.
- Results obtained by the matching experiments varied scene to scene even though majority of the tested images appeared to sharper and possessing higher contrast with decreasing image size. Investigation on objective techniques for analysis and classification of scene content and characteristics is suggested to link variability in image appearance results with original scene content.

Appendix A. Thumbnails of test images

A.1 16 ‘average’ scenes



A.2 Test images (in alphabetical order)



British museum 2



Couple



Florist 2



House



British museum



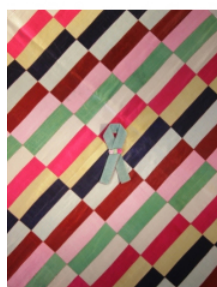
China plate



Florist



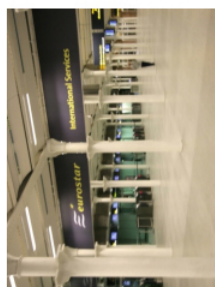
Halal shop



Bojagi



Carpet stall



Eurostar terminal



Girl



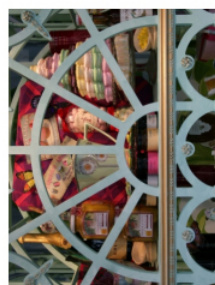
Bench



Butcher

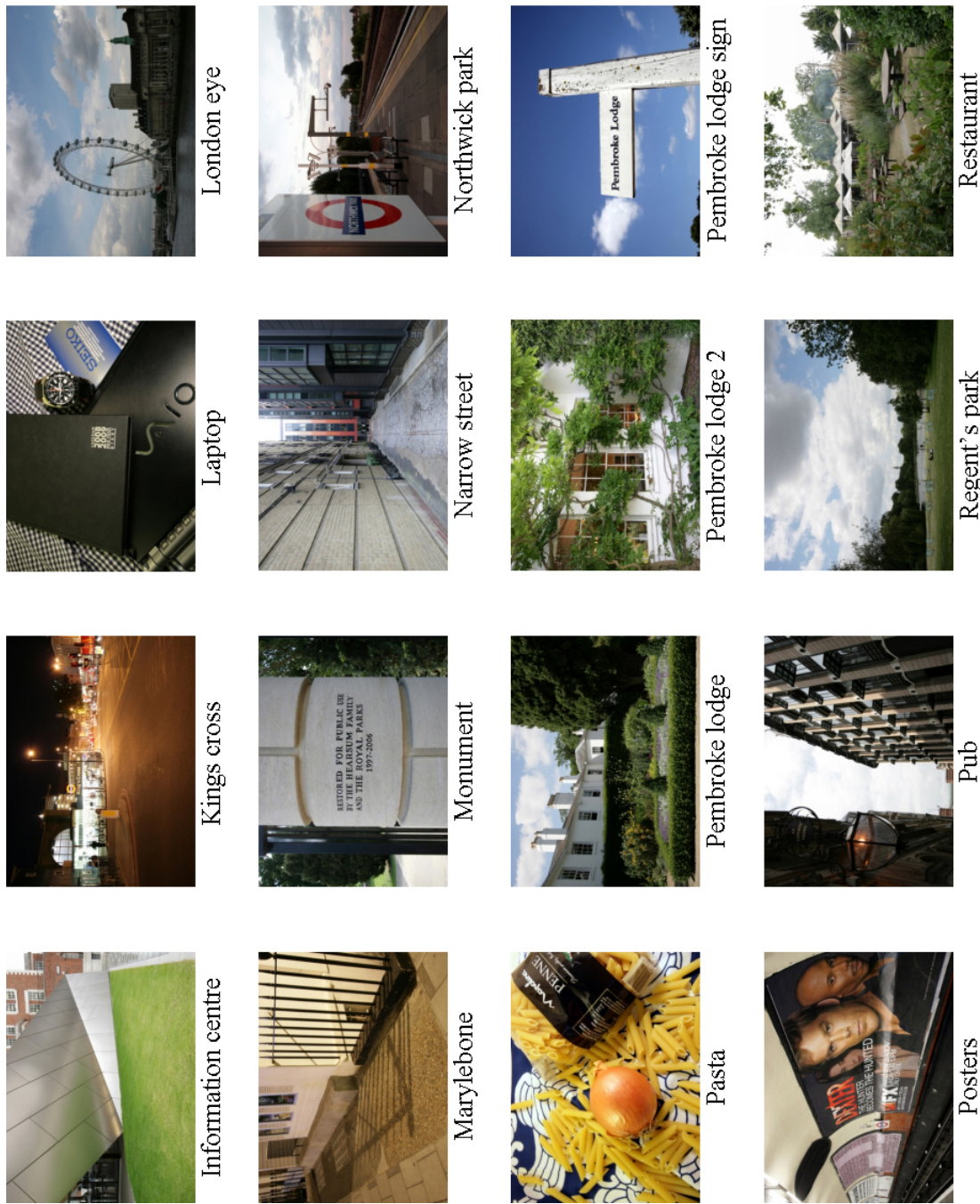


Docks

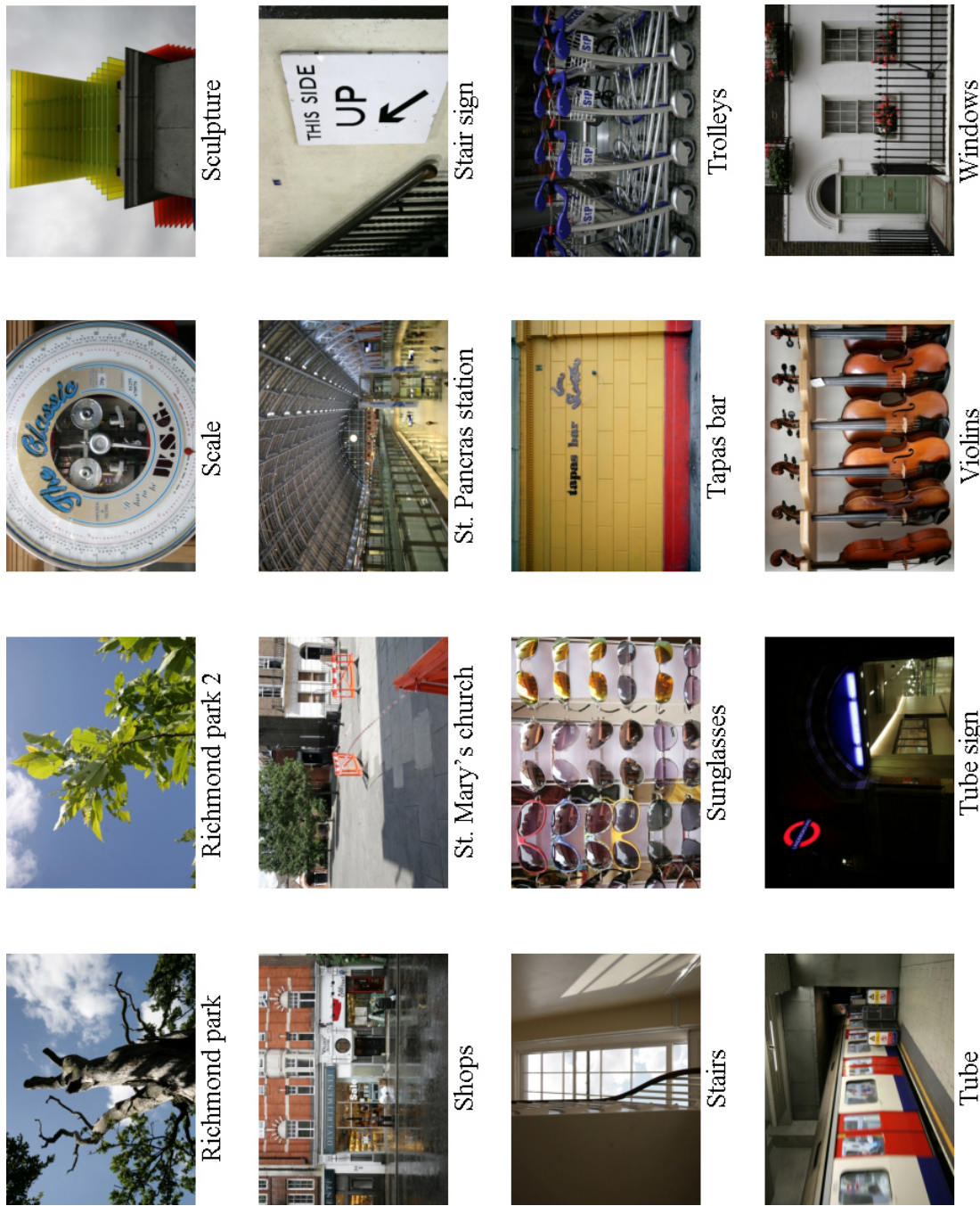


Fortnum & mason

A.2 Test images (in alphabetical order) - *continued*



A.2 Test images (in alphabetical order) - *continued*



Appendix B. Instructions for observers

The following instructions were provided to the observers before each psychophysical investigation.

B.1 Observer instructions for rank order experiments

Rank order experiments.

Instructions for the rank order test

Dear all,

Thank you very much for your contribution.

Before you start, please read the instructions below carefully. If you are not sure about some of the terms used in this test, feel free to ask me.

This test is based on the subjective appearance changes that occur when images are viewed at different sizes. The same image will be displayed on a LCD display in two different sizes. No other alterations (e.g. changes in contrast, sharpness, etc.) have been introduced to the images.

In achromatic mode, you will be asked to rank the attributes from 4 to 1 dependent on your impression of the images. The most affected attribute should be ranked as the maximum rank value of 4 and the least affected attribute as 1. They must be ranked **uniquely**.

There are no right or wrong answers as the test is based on subjective impression.

There are four sessions in total and each session with 64 image pairs will take approximately 30 to 45 minutes for achromatic images and slightly longer for the chromatic ones.

You are requested to start by selecting the **Apple iPhone** and **Achromatic** options for the first sitting, then the Canon 30D and Achromatic options for the second sitting. The third and fourth sittings should use the Apple iPhone and Chromatic options, and the Canon 30D and Chromatic options respectively.

B.2 Observer instructions for sharpness matching experiments

Sharpness matching.

Instructions for the sharpness matching test.

Dear all,

Thank you very much for your contribution.

Before you start, please read the instructions below carefully.

This test is based on the subjective appearance changes that occur when images are viewed at different sizes. A pair of randomly selected identical scenes will be displayed side-by-side on a LCD display. The image pairs, possessing different image qualities, will be displayed simultaneously in two different sizes. The larger version images are your reference images and the smaller ones are your test images.

For each image pair, you will be asked to match the perceived sharpness of the smaller test image to the larger reference image using a slider. Adjusting the slider bar will simulate a sharpness adjustment upon the smaller test image.

There are four sessions in total, and each session with 64 image pairs will take approximately 30 to 60 minutes.

There is a chin rest provided to keep the viewing distance at 60cm for all observers. However, you may adjust the chin rest level slightly to achieve the most comfortable position.

There are no right or wrong answers as the test is based on subjective impression.

If you are not sure about some of the terms used in this test, feel free to ask me.

B.3 Observer instructions for contrast matching experiments

Contrast matching.

Instructions for the contrast matching test.

Dear all,

Thank you very much for your contribution.

Before you start, please read the instructions below carefully.

This test is based on the subjective appearance changes that occur when images are viewed at different sizes. A pair of randomly selected identical scenes will be displayed side-by-side on a LCD display. The image pairs, possessing different image qualities, will be displayed simultaneously in two different sizes. The larger version images are your reference images and the smaller ones are your test images.

For each image pair, you will be asked to match the perceived contrast of the smaller test image to the larger reference image using a slider. Adjusting the slider bar will simulate a contrast adjustment upon the smaller test image.

There are four sessions in total, and each session with 64 image pairs will take approximately 30 to 60 minutes.

There is a chin rest provided to keep the viewing distance at 60cm for all observers. However, you may adjust the chin rest level slightly to achieve the most comfortable position.

There are no right or wrong answers as the test is based on subjective impression.

If you are not sure about some of the terms used in this test, feel free to ask me.

B.4 Observer instructions for result validation experiments

Validation of results.

Instructions for the result validation by pair rating experiment.

Dear all,

Thank you very much for your contribution.

Before you start, please read the instructions below carefully.

This test is based on the subjective appearance changes that occur when images are viewed at different sizes.

A pair of randomly selected identical scenes will be displayed side-by-side on a LCD display. A total of 16 scenes have been selected for this result validation test. For each reference image, a total of 3 image pairs have been prepared, consisting of 2 appearance matched pairs and 1 unaltered pair.

For each image pair, you will be asked to rate the pair in terms of appearance matching (from 10 being the most matching to 1 being the least matching). For the very first image pair, please avoid using the highest or the lowest ratings on this scale.

The test will take approximately 10 to 20 minutes.

There is a chin rest provided to keep the viewing distance at 60cm for all observers. However, you may adjust the chin rest level slightly to achieve the most comfortable position.

There are no right or wrong answers as the test is based on subjective impression.

If you are not sure about some of the terms used in this test, feel free to ask me.

B.5 Observer instructions for step validation experiments

Step validation.

Instructions for the step validation test.

Dear all,

Thank you very much for your contribution.

Before you start, please read the instructions below carefully.

A pair of randomly selected identical scenes will be displayed side-by-side on a LCD display. Both images are of the same scene, but possess different sharpness/contrast levels. A total of 64 scenes will be used for the step validation tests. For each reference image, a total of 12 test images possessing different sharpness/contrast levels have been generated.

For each image pair, you will be asked to select the image that you perceive to be sharper (or higher contrast for the contrast ruler images).

Each session will take approximately 10 minutes.

There is a chin rest provided to keep the viewing distance at 60cm for all observers. However, you may adjust the chin rest level slightly to achieve the most comfortable position.

There are no right or wrong answers as the test is based on subjective impression.

If you are not sure about some of the terms used in this test, feel free to ask me.

Appendix C. Publications

The following related papers were produced by the author during the production of this work and are reproduced in the following appendix.

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., “Identification of image attributes that are most affected with changes in displayed image size”, Proc. SPIE Image Quality and System Performance VI, 7242, 18-22 January 2009, San Jose, USA

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., Gupta, G., “Evaluation of perceived image sharpness with changes in the displayed image size”, Proc. SPIE Image Quality and System Performance IX, 8293, 22-26 January 2012, San Francisco, USA

Park, J. Y., Triantaphillidou, S., Jacobson, R. E., “Just noticeable differences in perceived image contrast with changes in the displayed image size”, Proc. SPIE Image Quality and System Performance XI, 9016, 2-6 February 2014, San Francisco, USA

Identification of image attributes that are most affected with changes in displayed image size

Jae Young Park, Sophie Triantaphillidou and Ralph E. Jacobson
University of Westminster, Watford Road, Harrow, HA1 3PT, UK

ABSTRACT

This paper describes an investigation of changes in image appearance when images are viewed at different image sizes on a high-end LCD device. Two digital image capturing devices of different overall image quality were used for recording identical natural scenes with a variety of pictorial contents. From each capturing device, a total of sixty four captured scenes, including architecture, nature, portraits, still and moving objects and artworks under various illumination conditions and recorded noise level were selected. The test set included some images where camera shake was purposefully introduced. An achromatic version of the image set that contained only lightness information was obtained by processing the captured images in CIELAB space. Rank order experiments were carried out to determine which image attribute(s) were most affected when the displayed image size was altered. These evaluations were carried out for both chromatic and achromatic versions of the stimuli. For the achromatic stimuli, attributes such as contrast, brightness, sharpness and noisiness were rank-ordered by the observers in terms of the degree of change. The same attributes, as well as hue and colourfulness, were investigated for the chromatic versions of the stimuli. Results showed that sharpness and contrast were the two most affected attributes with changes in displayed image size. The ranking of the remaining attributes varied with image content and illumination conditions. Further, experiments were carried out to link original scene content to the attributes that changed mostly with changes in image size.

Keywords: Image appearance, image quality, image size, scene content, scene analysis

1. INTRODUCTION

Various display technologies are available in mobile devices (e.g. digital camera, mobile phones, etc), yet LCDs are the most common displays in mobile devices with digital image capturing functions. Since often the quality of images on small LCD devices is satisfactory, the users check the captured images and tend to believe that the images they see on the small display is what they will get when their images will be displayed on larger displays. However, the images displayed on small display devices do not give full information, since image quality is highly dependent on the physical properties of the images [1]. Due to the properties and sizes of the built-in LCDs in capturing devices, the physical properties of the images maybe affected and give distorted information.

Image quality involves the assessment of many perceptual attributes such as image contrast, resolution, sharpness, graininess and colour [1]. Until recently, studies on the perceived quality of display devices were based on physical measurements, carried out to investigate one or more quality attributes rather than the overall visual appearance of the displayed images [2, 3, 4, 5]. Images are increasingly displayed and viewed in various sizes, under various viewing conditions. The subjective impressions of image quality will thus be affected by changes in image size [6], scene content (or complexity) [1] and also by environmental conditions, such as background colours and the intensity and colour of surrounds [7, 8]. Nevertheless, traditionally display characterisation techniques do not account for the displayed image size, or scene content.

In recent years there has been several studies investigating the effect of displayed image size on colour appearance and/or image quality [7, 8, 9]. In colour appearance investigations by Choi *et al*, the studies were mainly focused on changes of the size of uniform patches, as well as illumination level, surround and relative display luminance. Measurements in these studies were restricted to a certain area of the display, with a small coverage under certain viewing conditions. Further, Nezamabadi *et al* [9] investigated the relationship between changes in image size and

Image Quality and System Performance VI, edited by Susan P. Farnand, Frans Gaykema,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7242, 724210 · © 2009 SPIE-IS&T
CCC code: 0277-786X/09/\$18 · doi: 10.1117/12.805844

SPIE-IS&T/ Vol. 7242 724210-1

perceived contrast, using contrast matching techniques and artificially generated noise patterns of different spatial frequency content. In another study, Nezamabadi *et al* [8] investigated changes in perceived lightness and chroma with changes in visual angle and thus perceived image size. Spatial effects such as sharpness, noisiness and most importantly the appearance of digital image artefacts caused by varying image size, image content and illumination conditions have not been considered in depth. The work reported here has been carried out to identify those image attributes (both colour and spatial) that are most affected with changes in displayed image size. The experiment presented in this paper is part of a larger project which seeks to identify the degrees of change in image appearance by varying image size, pictorial content, whilst looking not only colour but further into spatial image quality attributes such as sharpness, noisiness and the appearance of digital image artefacts.

2. TEST STIMULI PREPARATION

2.1 Image capture

Two digital image capturing devices (with built-in LCDs) of different overall image quality were selected. An eight megapixel Canon 30D digital SLR camera, equipped with an EF-S10-22mm lens and a two megapixel Apple iPhone camera with a built-in lens were used for capturing the test images. For each captured scene both devices were set to the same focal length to record identical image frames. Because it was difficult to adjust shutter speed and aperture on the camera of Apple iPhone, several exposures were made with the Canon 30D in manual mode to match the captured image by the Apple iPhone camera. The most similarly exposed images by both cameras were selected by inspection to create an appropriate test set.

2.2 Image selection

Images were carefully selected to cover various possible situations and conditions captured by ordinary users. For each capturing device, a total of sixty four captured scenes including architectures, nature, portraits, still and moving objects and artworks under various illumination conditions with differing scene contrast and recorded noise levels were selected. The test sets also included some images where camera shake was purposefully introduced.

2.3 Image processing

The original captured images were too large to be displayed in full resolutions on the EIZO ColorEdge CG210 21.3" LCD which was used in this investigation. Thus, the test images were resized from their original sizes to 744 (v) x 560 (h) pixels using bi-cubic interpolation. This size is approximately the half of the LCD's horizontal resolution and is an appropriate image size to be displayed on commonly used XGA displays. In order to obtain the achromatic versions of the test stimuli, all resized images were then converted from RGB to CIELAB space and the lightness channel (L^*) was selected for the purpose.

3. PSYCHOPHYSICAL INVESTIGATION

3.1 Calibration and settings of the system

An EIZO ColorEdge CG210 LCD, driven by a Sony VAIO VGN-T92S computer with an on-board graphics controller was used in the psychophysical investigation. The LCD has a native resolution of 1,600 x 1,200 pixels (24 bits). The system was set to a white point luminance of 134.3 cd/m², a gamma of 2.2 and a colour temperature of 6500K, using the GretagMacbeth Eye-One Pro with Profilemaker5.

3.2 Software preparation and interface design

The application employed in this rank order psychophysical experiment was written in JavaScript. It was tested and optimised in Mozilla Firefox v3.0.1 web browser. A mid-grey background colour at a gamma of 2.2 (pixel value of 186 for all R, G and B channels) was selected. The application gathers personal information provided by the viewers before the experiments starts. During the experiment, each test image is displayed simultaneously in two different sizes, one at the original size (i.e. 744 (v) x 560 (h)) and the other equivalent to the size of the built-in LCDs in each capturing device. Test images are displayed in random order and in random positions, one on the left side and the other on the right side of the display (i.e. left: large image, right: small image OR the opposite). The application automatically writes the observation data and saves them as a text file. The display interface is illustrated in Figure 1.

3.3 Rank order method

The display device was placed in a totally dark environment to prevent viewing flare. Rank order tests were conducted by displaying the same image at two different displayed sizes, side by side (Figure 1). Each observer took the test four times: for each camera, they judged the achromatic versions and the chromatic versions of the test stimuli.

Observers were placed 60 cm away from the display and were asked to adjust their observation distance, if necessary. They were asked to rank-order the attributes that were affected with changes in displayed image size. For the achromatic version of the stimuli they ranked the following attributes (from 1, the most affected to 4, the least affected): contrast, brightness, sharpness and noisiness. For the chromatic versions, in addition to the previous attributes, they also rank-ordered hue and colourfulness (rank order from 1 to 6). A total of 19 observers, 9 females and 8 males took the experiment. Their age ranged between 20 and 60 years old. They had mostly imaging and design backgrounds.

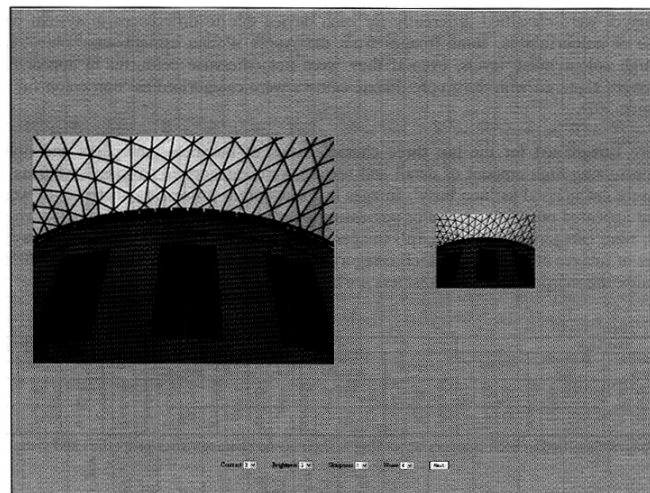


Figure 1. Interface for the psychophysical test page in achromatic mode.

4. RESULTS

4.1 Classification of test images

The original full size versions of the test stimuli were categorised by both an objective method suggested by Triantaphillidou *et al* [1] and visual inspection. Triantaphillidou *et al* have suggested a method to analyse and classify scenes by deriving certain scene metric values that identify how much or how little of a 'scene characteristic' they may process. To discuss the proximity of the scene metric values between the different scenes and therefore the similarity in their characteristics, they classified these scene metrics into four ranges, ordered by relative distance from the median. Values found to be within one sigma from the median value were considered 'average' and were split into 'average-to-low', if below the median and 'average-to-high', if above the median. Another two categories included 'extreme' values, i.e. values that were more than plus or minus one standard deviation away from the median, comprising the 'very low' category if more than one sigma below the median, and 'very high' if more than one sigma above the median.

To investigate the relationships between the attribute ranking and the test-scene content, the test stimuli in this work were categorised according to five different image characteristics: three characteristics were related to both scene content and system performance and two characteristics were related only to the system performance. For examining the lightness (i.e. how light/dark they may be) and the colourfulness (i.e. how colourful/non-colourful they may be) characteristics of the test images, the original stimuli were categorised by both the suggested objective scene analysis method and visual inspection. For examining the busyness, sharpness and noisiness of the original test stimuli, only visual inspection in a dark surround was used.

Table 1 shows the selected image characteristics and the number of images that fall in each category, for both the Canon 30D and the Apple iPhone cameras. For the 'light/dark' characteristic, the test images obtained by both devices were similarly classified – similar number of images in each category. However, for the 'colourful/non-colourful' characteristic, the images were classified differently for each device due to differences in system performances. For example, in the case of colourfulness, some images from the Apple iPhone camera were objectively classified as 'colourful' due to high colour noise levels, even if they were not otherwise colourful in appearance. Also, more 'moderately dark' images captured with the Apple iPhone camera were categorised as 'non colourful' than when these captured with the Canon 30D.

The test images were categorised for the last three characteristics of busyness, sharpness and noisiness by visual inspection. Images possessing high amount of detail and texture were categorised as 'busy' whereas those possessing slow varying areas were categorised as 'non busy'. Images in-focus and/or possessing sharp edges were categorised as 'sharp' and those that appeared out-of-focus and/or possessing blurred and moving objects and/or where camera shake had been introduced, were categorised as 'un-sharp'. Images appeared to possess any visible noise were categorised as 'noisy'. The numbers of images classified into each category for the last two characteristics were different for the two camera systems, because the sharpness and the noisiness are highly dependent on system performance.

	Dark/Light				Colourful				Busy		Sharp		Noisy	
	Dark	Moderately dark	Moderately light	Light	Colourful	Moderately colourful	Moderately non colourful	Non colourful	Yes	No	Yes	No	Yes	No
Canon 30D	8	24	22	10	17	15	21	11	30	34	49	15	30	34
Apple iPhone	8	24	21	11	21	11	11	21			43	21	62	2

Table 1. Numbers of images classified in each category according to their scene characteristics/properties and visual inspection.

4.2 Results from attribute ranking

The ranked scale data for all test stimuli were averaged for each mode (achromatic and chromatic) and each camera. Tables 2 and 3 present the average ranks for each attribute and these ranks also being converted to an ordinal scale (from 1 to 4, or 1 to 6) that represents the most to the least affected attribute. For the achromatic version of the stimuli, the maximum average rank was 4. 6 was the maximum average rank for the chromatic version of the stimuli. Attributes with higher average ranks were more affected and those with lower average ranks.

Canon 30D			Achromatic				Chromatic					
			Contrast	Brightness	Sharpness	Noisiness	Contrast	Brightness	Sharpness	Noisiness	Hue	Colourfulness
Dark	Dark	Average	2.74	2.30	2.78	2.18	4.18	3.61	4.65	2.91	2.52	3.13
		Order	2	3	1	4	2	3	1	5	6	4
	Moderately dark	Average	2.73	2.24	3.05	1.98	4.10	3.64	4.66	2.42	2.65	3.52
		Order	2	3	1	4	2	3	1	6	5	4
	Moderately bright	Average	2.78	2.36	3.00	1.85	4.20	3.74	4.62	2.21	2.71	3.52
		Order	2	3	1	4	2	3	1	6	5	4
	Bright	Average	2.63	2.44	3.04	1.89	4.07	3.86	4.71	2.43	2.65	3.29
		Order	2	3	1	4	2	3	1	6	5	4
Colourful	Colourful	Average	2.66	2.40	3.03	1.92	4.00	3.75	4.53	2.24	2.62	3.87
		Order	2	3	1	4	2	4	1	6	5	3
	Moderately colourful	Average	2.76	2.20	2.97	2.08	4.16	3.62	4.66	2.53	2.55	3.48
		Order	2	3	1	4	2	3	1	6	5	4
	Moderately non colourful	Average	2.74	2.32	3.01	1.93	4.17	3.73	4.72	2.43	2.73	3.22
		Order	2	3	1	4	2	3	1	6	5	4
	Non colourful	Average	2.82	2.38	2.96	1.84	4.28	3.70	4.70	2.48	2.72	3.12
		Order	2	3	1	4	2	3	1	6	5	4
Busy	Yes	Average	2.76	2.31	3.03	1.90	4.11	3.63	4.69	2.33	2.67	3.58
		Order	2	3	1	4	2	3	1	6	5	4
	No	Average	2.71	2.33	2.97	1.99	4.17	3.77	4.63	2.48	2.65	3.31
		Order	2	3	1	4	2	3	1	6	5	4
Sharp	Yes	Average	2.76	2.36	2.95	1.93	4.17	3.69	4.63	2.38	2.66	3.48
		Order	2	3	1	4	2	3	1	6	5	4
	No	Average	2.67	2.21	3.14	1.99	4.06	3.73	4.74	2.53	2.66	3.28
		Order	2	3	1	4	2	3	1	6	5	4
Noisy	Yes	Average	2.67	2.27	3.01	2.05	4.09	3.65	4.68	2.62	2.57	3.39
		Order	2	3	1	4	2	3	1	5	6	4
	No	Average	2.79	2.37	2.99	1.86	4.19	3.75	4.63	2.22	2.74	3.48
		Order	2	3	1	4	2	3	1	6	5	4
Average ranks from all stimuli			2.74	2.32	3.00	1.95	4.14	3.70	4.65	2.41	2.66	3.43
Overall Order			2	3	1	4	2	3	1	6	5	4

Table 2. Average ranks and rank order of the image attributes from the image stimuli obtained by the Canon 30D.

Apple iPhone			Achromatic			Chromatic		
	Average	Order	Contrast	Brightness	Noisiness	Contrast	Brightness	Noisiness
Dark	Dark	Average	2.75	2.16	2.52	2.57	3.52	3.23
		Order	1	4	3	2	3	4
	Moderately dark	Average	2.64	2.26	2.95	2.15	3.71	3.43
		Order	2	3	1	4	2	3
Moderately bright	Moderately bright	Average	2.69	2.17	3.04	2.10	3.51	3.29
		Order	2	3	1	4	2	3
	Bright	Average	2.59	2.18	3.02	2.21	3.41	3.20
		Order	2	4	1	3	3	5
Colourful	Colourful	Average	2.70	2.13	3.03	2.14	3.47	3.39
		Order	2	4	1	3	3	4
	Moderately colourful	Average	2.67	2.24	2.94	2.16	3.71	3.33
		Order	2	3	1	4	2	3
Non colourful	Moderately non colourful	Average	2.70	2.12	2.90	2.28	3.59	3.35
		Order	2	4	1	3	3	4
	Non colourful	Average	2.61	2.30	2.87	2.23	3.59	3.22
		Order	2	3	1	4	3	4
Busy	Yes	Average	2.70	2.19	3.05	2.05	3.63	3.47
		Order	2	3	1	4	2	3
	No	Average	2.63	2.21	2.84	2.32	3.51	3.18
		Order	2	4	1	3	3	5
Sharp	Yes	Average	2.70	2.22	2.92	2.16	3.57	3.33
		Order	2	3	1	4	2	3
	No	Average	2.60	2.17	2.97	2.26	3.57	3.29
		Order	2	4	1	3	3	5
Noisy	Yes	Average	2.66	2.20	2.94	2.21	3.55	3.31
		Order	2	4	1	3	2	5
	No	Average	2.91	2.29	2.94	1.85	4.22	3.59
		Order	2	3	1	4	2	4
Average ranks from all stimuli			2.66	2.20	2.94	2.19	3.57	3.32
Overall Order			2	3	1	4	2	5

Table 3. Average ranks and rank order of the image attributes from the image stimuli obtained by the Apple iPhone camera.

5. DISCUSSION

Figure 2 shows the average ranks for all images captured by both cameras. The higher the ranking the more affected the attribute.

Overall, for the achromatic version of the stimuli obtained by both cameras, observers ranked sharpness as the most affected and contrast as the second most affected attributes with changes in the displayed image size. Brightness and noisiness were ranked as the third and fourth, respectively. For the stimuli obtained by the Apple iPhone camera, the observers also ranked brightness and noisiness as third and fourth respectively, but with a smaller difference in average ranks (2.19 and 2.20).

For the chromatic version of the stimuli captured with the Canon 30D, observers again ranked sharpness, contrast, and brightness in the same order as for the achromatic stimuli. Colourfulness and hue were found to be affected less, whilst noisiness was the least affected.

For the chromatic version of stimuli from the Apple iPhone camera, observers again ranked sharpness and contrast as the first two most affected attributes. However, noisiness, colourfulness and brightness were ranked as third, fourth and fifth whilst hue was ranked last.

The ranking of the attributes for the chromatic version of stimuli from the Apple iPhone camera was different, with noise and colourfulness having higher ranks, compared to the Canon 30D. This result is related to the relatively lower average lightness and the higher level of chromatic noise that were presented in most of the test stimuli originating from the Apple iPhone camera.

Overall rank-orders of the attributes are as below:

Achromatic stimuli:	Sharpness > Contrast > Brightness > Noisiness	(Both devices)
Chromatic stimuli:	Sharpness > Contrast > Brightness > Colourfulness > Hue > Noisiness	(Canon 30D)
	Sharpness > Contrast > Noisiness > Colourfulness > Brightness > Hue	(Apple iPhone)

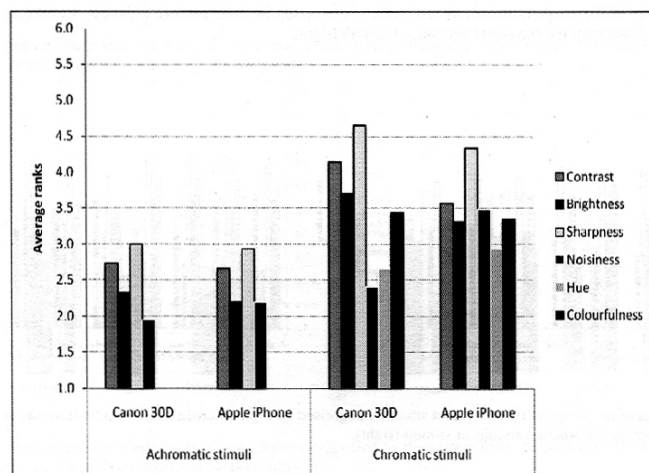


Figure 2. Average ranks from all test stimuli.

Figures 3 to 7 present the average ranks in relation to image categories that were described in section 4.1.

In general, test stimuli categorised as 'dark', thus contain higher level of noise, sharpness was affected less and noisiness was affected more compared with other stimuli classified in different categories (figure 3). For test stimuli obtained by the Canon 30D, the average ranks for brightness was slightly increased as the average lightness of the test stimuli increased (figure 3). In a similar fashion, the average ranks of colourfulness was increased with 'colourful' stimuli and decreased with 'non colourful' stimuli (figure 4).

Average ranks of noisiness were decreased with the stimuli categorised as 'busy' whilst that with stimuli categorised as 'non busy' they were increased (figure 5). This is because of the noise being masked by the high frequency information in the scene. Average ranks of contrast were increased when stimuli categorised as 'sharp', whilst that of sharpness and noisiness were decreased (figure 6). The noise level of the test stimuli were also related to hue and colourfulness in the chromatic versions (figure 7). Average ranks of noisiness were affected more with stimuli categorised as 'noisy', indicating that when noise is visible in the image then noisiness becomes an important attribute with respect to changes in image size (figure 7).

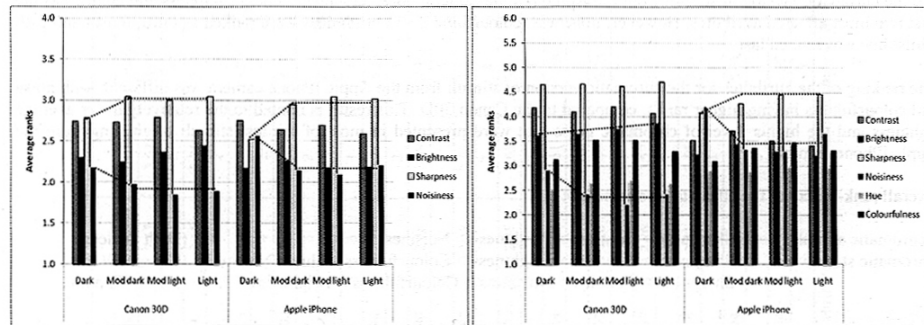


Figure 3. Average ranks of the image attributes of test stimuli categorised by their average lightness. Results from the achromatic versions of stimuli (left) and from the chromatic versions of stimuli (right).

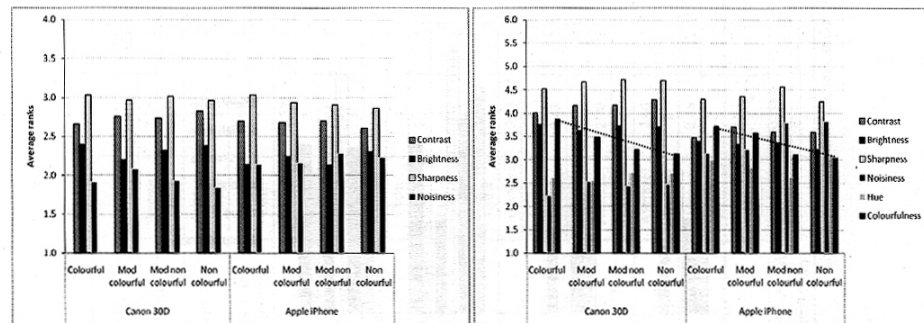


Figure 4. Average ranks of the image attributes of test stimuli categorised by their colourfulness. Results from the achromatic versions of stimuli (left) and from the chromatic versions of stimuli (right).

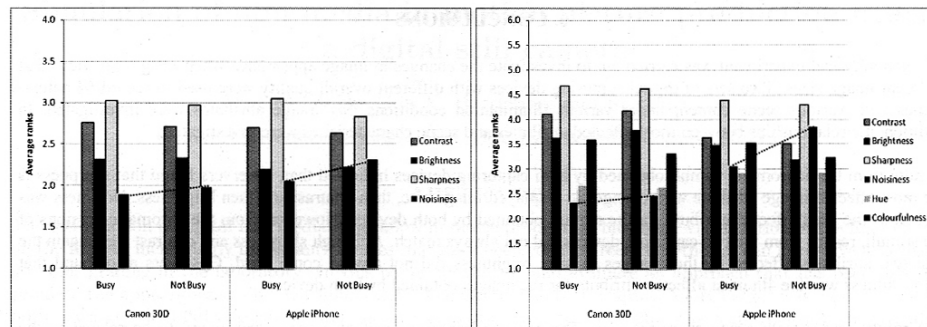


Figure 5. Average ranks of the image attributes of test stimuli categorised by their busyness. Results from the achromatic versions of stimuli (left) and from the chromatic versions of stimuli (right).

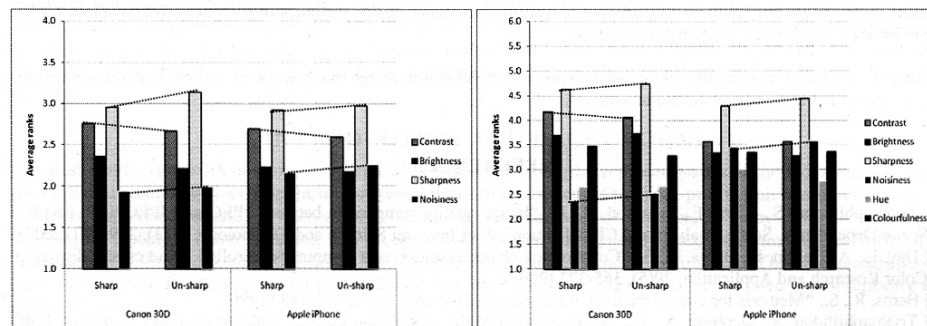


Figure 6. Average ranks of the image attributes of test stimuli categorised by their sharpness. Results from the achromatic versions of stimuli (left) and from the chromatic versions of stimuli (right).

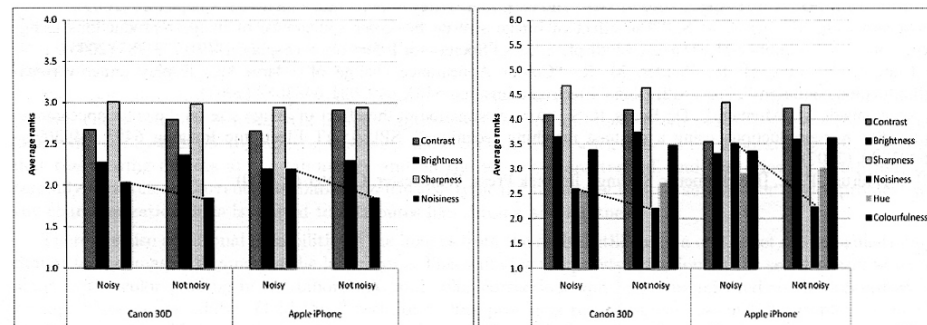


Figure 7. Average ranks of the image attributes of test stimuli categorised by their noise level. Results from the achromatic versions of stimuli (left) and from the chromatic versions of stimuli (right).

6. CONCLUSIONS

A psychophysical experiment was carried out to investigate the changes in image appearance when images are viewed at different image sizes. Two sets of digital capturing devices with different overall quality were used to record 64 natural scenes with various scene content under various illumination conditions. Six image attributes were investigated. In addition, the relationships between most affected attributes and scene characteristics were investigated.

Results from the achromatic stimuli obtained by both capturing devices indicated that, observers found that sharpness is the most affected image attribute with changes in displayed image size, then contrast and then brightness. Noisiness was ranked as the least affected attribute for the stimuli obtained by both devices. However, from the chromatic versions of the stimuli, results from the two capturing devices did not always match. Although sharpness and contrast were again the first two attributes affected by the changes in size, brightness did not always come third. Observers also found that colourfulness was the 4th most affected attribute for the images obtained by both devices.

The results varied with scene characteristics. The average lightness level of a scene was found to be related to the attributes of sharpness and noisiness, whilst the colourfulness of a scene was found to be related more to the attribute of colourfulness. The busyness of a scene was found to be also related to noisiness, i.e. as busyness increased noisiness became less and less important. The sharpness of a scene was related to the attributes of sharpness and noisiness whilst the noise level of a scene was related to noisiness.

Further research work is currently taking place to model the changes in image sharpness with changes in displayed image size using various sharpness matching techniques.

REFERENCES

- [1] Triantaphillidou, S., Allen, E., Jacobson, R., E., "Image quality comparison between JPEG and JPEG2000, Part II- Scene Dependency, Scene Analysis and Classification", J. of Imaging Science and Technology, 51(3), 259-270 (2007).
- [2] Day, E., A., Taplin, L., Berns, R., S., "Colorimetric characterization of a computer-controlled liquid crystal display", Color Research and Application, 29(5), 365-373 (2004).
- [3] Berns, R., S., "Methods for characterizing CRT displays", Displays, 16(4), 173-182 (1996).
- [4] Triantaphillidou, S., Jacobson, R., E., "Measurements of the modulation transfer function of image displays", J. of Imaging Science and Technology, 48 (1), 58-65 (2004).
- [5] MacDonald, L. W. and Luo, M. R., [Colour Imaging: Vision and Technology], John Wiley & Sons Ltd, Chichester, Chapter 15, (1999).
- [6] Fairchild, M. D., [Color Appearance Models, 2nd ed.], John Wiley & Sons Ltd, Chichester, Chapter 20 (2005).
- [7] Nezamabadi, M., Berns, R. S., "The effect of image size on the color appearance of image reproductions using colorimetrically calibrated LCD and DLP displays", J. of Society for Information Display, 14(9), 773-783 (2005).
- [8] Choi, S. Y., Luo, M. R., Pointer, M. R., "Colour Appearance change of a large size display under various illumination conditions", Proc. SPIE/IS&T Electronic Imaging 6493, 6493081-6483089 (2007).
- [9] Nezamabadi, M., Montag, E. D., Berns, R. S., "An investigation of the effect of image size on the color appearance of softcopy reproductions using a contrast matching technique", SPIE/IS&T Electronic Imaging 6493, 6493091-6493099 (2007).
- [10] Engeldrum, P. G., [Psychometric Scaling], Imcotek Press, Winchester, Chapter 6 (2000).

Evaluation of perceived image sharpness with changes in the displayed image size

Jae Young Park, Sophie Triantaphillidou, Ralph E. Jacobson and Gaurav Gupta
University of Westminster, Watford Road, Harrow, HA1 3TP, UK

ABSTRACT

In this paper an evaluation of the degree of change in the perceived image sharpness with changes in displayed image size was carried out. This was achieved by collecting data from three psychophysical investigations that used techniques to match the perceived sharpness of displayed images of three different sizes. The paper first describes a method employed to create a series of frequency domain filters for sharpening and blurring. The filters were designed to achieve one just-noticeable-difference (JND) in quality between images viewed from a certain distance and having a certain displayed image size (and thus angle of subtense). During psychophysical experiments, the filtered images were used as a test series for sharpness matching. For the capture of test-images, a digital SLR camera with a quality zoom lens was used for recording natural scenes with varying scene content, under various illumination conditions. For the psychophysical investigation, a total of sixty-four original test-images were selected and resized, using bi-cubic interpolation, to three different image sizes, representing typical displayed sizes. Results showed that the degree of change in sharpness between images of different sizes varied with scene content.

Keywords: Image appearance, image quality, image fidelity, image size, sharpness matching, scene content

1. INTRODUCTION

Digital images are increasingly viewed in various sizes in different display devices, such as tablet, mobile phone, laptop and desktop monitors. Image appearance, however, is a phenomenon of visual perception and thus it is naturally affected by several factors including the viewing conditions. Changes in image size, or the viewing distance have been reported to lead to changes in various aspects of image appearance [1, 2, 3, 4, 5].

A study concerning the identification of image attributes that are most affected by changes in the displayed image size was previously carried out by the authors [6]. It considered various image attributes, including both spatial and color and identified sharpness and contrast to be the two most affected attributes by changes in the displayed image size, and hue the least affected attribute. Similar results were found in a recent study conducted by Wang *et al* [7].

The main goals of our research are to quantify and model the perceived changes in individual image attributes with changes in the displayed image size and use these to match final image appearance at various display sizes. In this paper, we specifically focus on the quantification of changes in perceived sharpness with respect to changes in displayed image size. The softcopy ruler method (ISO 20462-3 [8]) has been adapted to create test series of images that vary in sharpness and are spaced by equal sharpness intervals.

2. TEST STIMULI PREPARATION

A large number of test image images were captured, using a Canon 30D digital SLR camera, equipped with a CMOS sensor of 3504(*h*) x 2336(*v*) pixel resolution and an EF-S10-22mm zoom lens that provide a 35mm equivalent focal length of 16-35mm. A fixed focal length of 22mm was used for capturing images at different ISO settings and lens apertures. A total of sixty-four captured scenes were selected after visual inspection of quality and scene content. The selected scenes included architectural and natural scenes, portraits, artworks, still and moving objects. They were recorded under various illumination conditions, had differing original scene contrast and recorded noise levels, various amounts of fine detail, strong lines and edges. Camera shake was purposefully introduced to some images in the test set to investigate its appearance at different image sizes. The original images were resized, using bi-cubic interpolation, to obtain three versions of the same scenes of different sizes. The test image dimensions were 744(*h*) x 560(*v*) pixels, 372(*h*)

Image Quality and System Performance IX, edited by Frans Gaykema, Peter D. Burns,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8293, 82930J · © 2012 SPIE-IS&T
CCC code: 0277-786X/12/\$18 · doi: 10.1117/12.908201

SPIE-IS&T/ Vol. 8293 82930J-1

x 280(v), and 186(h) x 140(v) and represented large, medium and small sizes commonly displayed on computer and mobile device monitors. The small size was based on prevalent pixel dimensions of LCD of DSLR capturing devices. The large size was approximately half of the EIZO ColorEdge CG210 LCD's native horizontal resolution, which was later used for image appearance matching.

A series of images with different sharpness levels were then created for each resized version (small, medium, and large). The softcopy ruler method was adapted to generate sharpening and blurring filters for the purpose. The point of these filters was to introduce systematic modifications in sharpness, while taking into account the performance of the imaging systems and fixed viewing distance. The following sections describe the steps implemented to create sets of images with varying sharpness, using frequency domain sharpening and blurring filters.

2.1 Modelling of the system MTF and relative quality JNDs

The softcopy ruler standard describes a methodology to generate image stimuli with varying sharpness, based on the imaging system performance, specifically the MTF of the system and its modifications. The implementation steps include: 1) measurement of the system's MTF, 2) conversion of the unit of spatial frequency from cycles/mm to cycles/degree to account for the viewing distance/visual angle, 3) determination of a reciprocal measure of the system's bandwidth in cycles/degree, k , which is used for the evaluation of relative quality JND for the specific imaging system, 4) evaluation of relative quality JND.

The standard presents Equations (1) and (2) to evaluate the quality of a reference stimuli generated by an imaging system and based on that reference to generate softcopy ruler stimuli.

Equation (1) represents the monochromatic MTF of an on-axis diffraction limited lens, $m(v)$.

$$\begin{aligned} m(v) &= \frac{2}{\pi} (\cos^{-1} kv - kv\sqrt{1 - (kv)^2}) & \text{when } kv \leq 1 \\ m(v) &= 0 & \text{when } kv > 1 \end{aligned} \quad (1)$$

where v is the spatial frequency in cycles/degree subtended at the eye of the observer and k is a reciprocal measure of the system bandwidth in cycles/degree.

Equation (2) is used to compute the relative quality JNDs of the system associated with given k values.

$$JNDs = \frac{17249 + 203793k - 114950k^2 - 3571075k^3}{578 - 1304k + 357372k^2} \quad (0.01 \leq k \leq 0.26) \quad (2)$$

The relative quality scale in the standard has been calibrated in terms of multivariate JND increments, which are larger than univariate increments, i.e. increments varying in one attribute, such sharpness only. In the case of sharpness, Keelan has found that JND increments of quality to be approximately twice as large as the JNDs increments found in sharpness [9].

2.2 Measurement of the Opto-Electronic Conversion Function (OECF [10]) and the Electro-Optic Conversion Function (EOCF)

Our experiment involved the measurement of the transfer functions of both capturing and display devices and the determination of the system gamma, which was further used for signal linearization during SFR measurements (Section 2.3). The transfer functions for both devices were measured individually and in combination. A Kodak Q-13 20-step grayscale was used for the purpose.

First a total of twelve frames were captured with the Canon 30D, at four different ISO camera settings (100, 400, 800, and 1600) and using three different apertures ($f/4.5$, $f/8.0$, and $f/16$) - i.e. ISO and aperture settings employed in test image capture. Normalized average output pixel values were plotted against normalized input luminances, as shown in Figure 1 (a). The captured images were then displayed on the calibrated EIZO LCD to a white point luminance of 120 cd/m², a gamma of 2.2, and a color temperature of D₆₅. The average output luminances were measured using a calibrated chroma meter (Konica Minolta CS-200). Normalized output luminance was plotted against normalized input pixel values, as

shown in Figure 1 (b). The tone reproduction of the combined system is illustrated in log-log space in Figure 1 (c). The overall gamma was derived from the linear part of the curve (mid-tones, where original SFR target luminances are falling, Section 2.3).

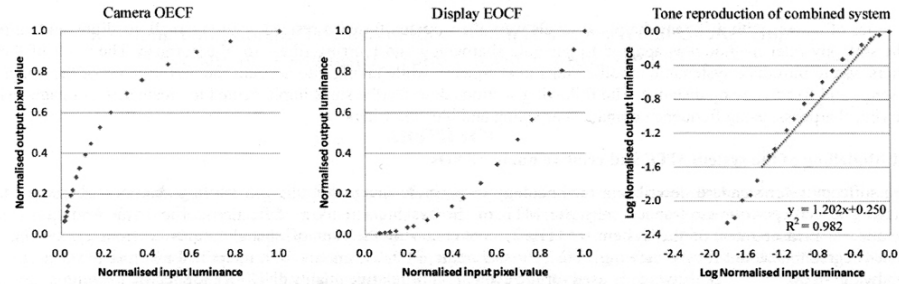


Figure 1. (a) OECF of the capturing device in linear space; (b) EOCF of the display device in linear space; (c) tone reproduction of the combined system in log-log space, with a gamma at the linear region.

2.3 Camera-display system MTF

The slanted edge method (ISO 12233:2000 [11]) was implemented to measure the combined spatial frequency response (SFR) of the capturing and display devices. A test target with SVG patterns (scalable vector graphics patterns, Figure 2 (a)) was created using Imatest® software [12]. The test target had a contrast ratio of 20:1, and a gamma of 2.2. It was displayed on the calibrated EIZO LCD and was captured using the Canon camera with the zoom lens set to focal length of 22mm and $f/8$, from a distance of 100cm. The overall system gamma (Section 2.2) was taken into account in the computation of the system SFR. All four edges of the central SVG patch were used. Horizontal and vertical SFRs were weighted equally to determine the 'average' system SFR, shown in Figure 2 (b). The spatial frequency was then converted from cycles/pixel to cycles/degree by taking account the observation distance, 60 cm, as shown in Figure 2 (c). This curve represents the measured system MTF in the following steps.

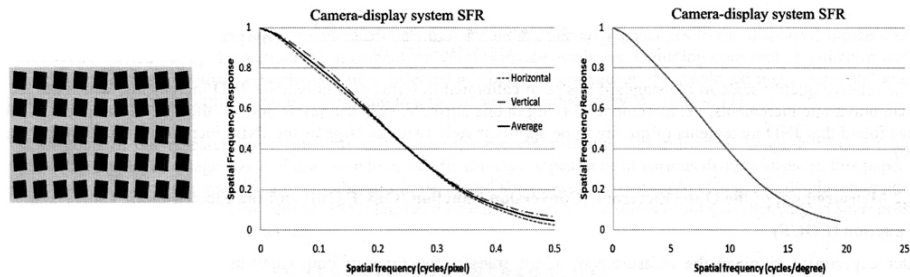


Figure 2. (a) SVG patterns created using Imatest® software; (b) the camera-display system SFR up to the Nyquist limit of the camera in cycles/pixel; (c) the camera-display system SFR in cycles/degree.

2.4 Determination of the reciprocal measure of the system bandwidth, k

A set of model curves was computed by varying k values in Equation (1). They are plotted in Figure 3 (b). The combined imaging system MTF was compared with the modelled curves and the closest model curve was selected. The k value, which represents the 'shape parameter' of the model curve was $k=0.0432$. The relative quality JND value associated with the system bandwidth k was then calculated for an average scene using Equation (2). The relative quality JND associated with $k=0.0432$ was found to be 21.5.

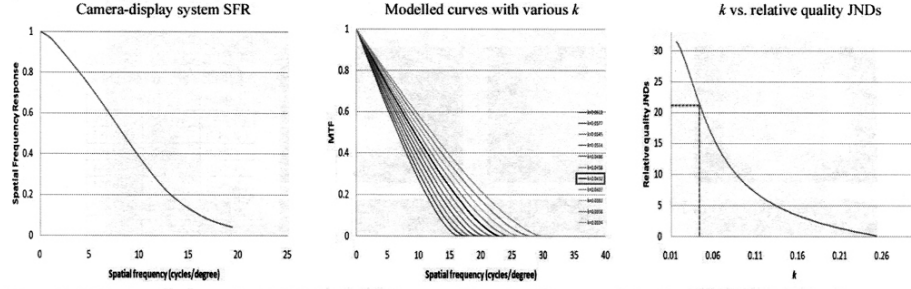


Figure 3. (a) Camera-display system MTF; (b) modeled curves with various k to approximate the sharpness of the imaging system; (c) relative quality JNDs vs. k .

2.5 Creation and application of frequency domain filters

Based on the relative quality JND of the imaging system, a total of thirteen k values associated with relative quality JND values, ranging between 21.5 and 9.5 were selected with a constant interval of 1 JND. The curves obtained using k values were then divided by the combined system MTF. Modulation transfer factors beyond the Nyquist frequency of the LCD display were replaced by 0. The resulting curves represented the functions to be used for blurring; they are illustrated in Figure 4 (b). Vertically flipped versions of these represented the functions for sharpening, shown in Figure 4 (c). The x-axes in Figures 4 (b) and (c) represent the image and filter dimensions. 1-D Gaussian filter functions [13], which represent the curves in Figure 4 (b), were obtained for the medium and small size images, using Oakdale Engineering Datafit Ver.9 [14]. Equation (3) represents the blurring and sharpening filters, H ,

$$\begin{aligned} H &= 1 - e^{-a \cdot b^D} && \text{for blurring} \\ H &= 1 + e^{-a \cdot b^D} && \text{for sharpening} \end{aligned} \quad (3)$$

where D is the digital image dimensions of the images and their spectra, and a , and b are the variables representing the sizes of the filter apertures.

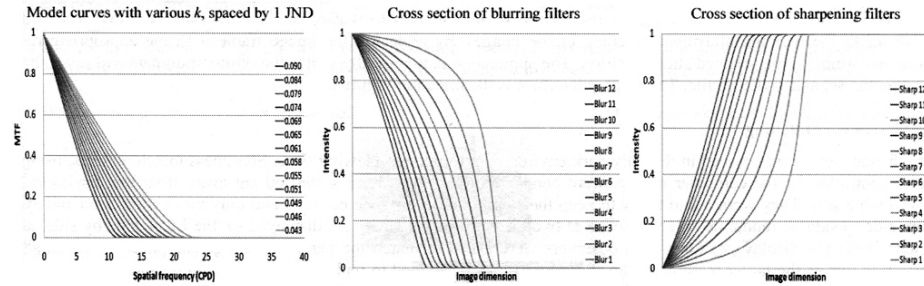


Figure 4. (a) Modeled curves with various k , spaced by 1 relative quality JND; (b) cross section of blurring and (c) sharpening filters, ranging from 1 to 12 relative quality JNDs.

Images of blurring filter functions were illustrated in Figure 5. The filtering operations were carried out using MATLAB. Images were converted from sRGB space to CIELAB space. The filters were applied to the lightness, L^* , channel only. The mean luminance component was subtracted during filtering and added back to avoid normalization of the luminance

channel. A total of 25 ruler images, possessing different perceived sharpness levels (original, 12 blurred versions and 12 sharpened versions) were generated for the small and medium size images.

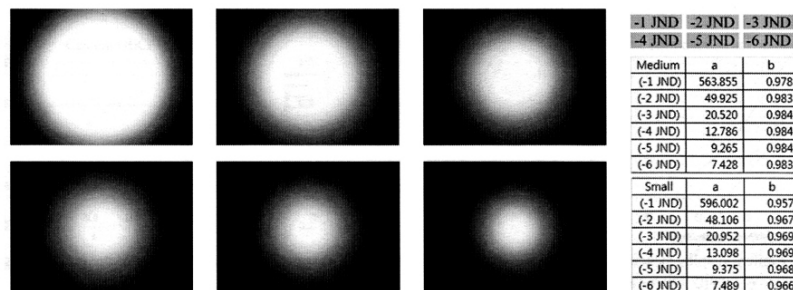


Figure 5. Images of blurring filters and the a and b values for the filter aperture.

3. PSYCHOPHYSICAL INVESTIGATION

3.1 Calibration and settings of the system

The EIZO ColoeEdge 210 21.3" LCD, driven by a Sony VAIO VGN-T92S computer with an on-board graphics controller, was used in the psychophysical investigation. The LCD has a native spatial resolution of 1,600 x 1,200 pixels and a tonal resolution of 24 bits. The system was set to a white point luminance of 120 cd/m², a gamma of 2.2 and a color temperature of D₆₅, using the GretagMacbeth Eye-One Pro with Profilemaker5.

3.2 Software preparation and interface design

The application employed in the sharpness matching experiment was written in PHP, HTML and CSS, the user interface being controlled using JavaScript. It was tested and optimized in Mozilla Firefox v5.0 web browser [15]. A mid-grey background (pixel value of 186 for all R, G and B channels at a gamma of 2.2) was selected. The application gathers some personal information provided by the observers before the experiments starts. The experiment consisted of three matching sharpness sessions: small image size to large image size, small image size to medium image size, and medium image size to large image size. Each test image was displayed simultaneously in two different sizes. The test images were displayed in random order and display sides, one on the left side and the other on the right side of the display. Observers were asked to match the sharpness of the smaller images to that of larger ones using a slider. The slider simulates to the user the blurring/sharpening of the images, by replacing the image frame with the appropriate ruler image according to the selected slider position. The application automatically writes the observation data and saves them in a comma separated (CSV) file. The display interface is illustrated in Figure 6.

3.3 Sharpness matching

The display device was placed in a totally dark environment to prevent viewing flare. Sharpness matching tests, using a slider controlled by the computer mouse, were conducted. Observers were seated 60 cm away from the display on a comfortable seat. They were asked to try to keep their observation distance constant and only move their head and eyes from side to side. A randomly selected scene at two different image sizes was displayed on the LCD, side by side, at a random left/right display positions. The observers were asked to match the perceived image sharpness of the smaller version to that of the larger reference. The slider bar was programmed to replace ruler images in response to changes in the slider position. The observers completed the experiment in three separate sessions. Each observation took less than one hour per session; one session was conducted per day to avoid fatigue. A total of twenty-two observers, with normal visual acuity, took the experiment. Their age ranged between 20 and 40 years old. Fifteen observers had imaging and design backgrounds.



Figure 6. Graphical user interface of the sharpness matching experiment.

4. RESULTS

4.1 Compensation of sharpness loss as a result of bi-cubic interpolation

The test stimuli were converted from original capture to three different sizes using bi-cubic interpolation. To quantify the effects of the resizing operation on the system MTF, two scenes possessing strong, measureable edges were selected by visual inspections. The SFRs of two different scenes were measured for all three sizes from a selected edge in each scene. A sharpness loss, as seen in Figure 7, was detected after the interpolation. The extent of this effect was however found to be scene dependent and is known to be very non-linear and thus limit the interpretation of the results. The sharper image, shown in Figure 7 (b) to have a generally higher SFR, was affected less by the interpolation, while the less sharp image, with a lower SFR shown in Figure 7 (a), was affected more. Average differences in sharpness due to interpolation were approximately 2.75 JNDs between small and medium sized images, also between medium and large sized images.

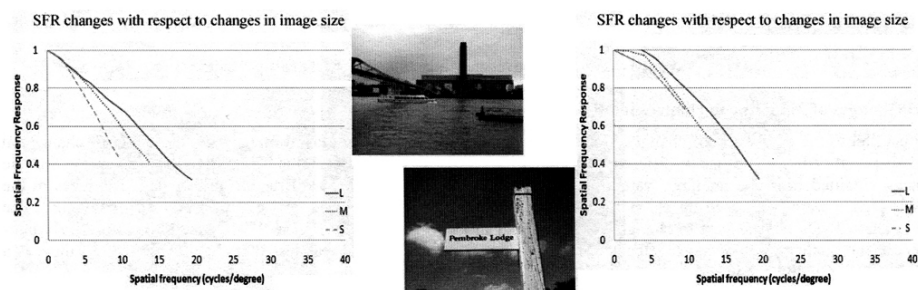


Figure 7. Effect of bi-cubic interpolation on the image sharpness with respect to image size.

4.2 Results from psychophysical tests

The raw data obtained from all observers were first adjusted by taking into account the sharpness losses occurring during the interpolation stage. Then the results were averaged for each individual scene and the standard deviation, σ , for each scene was computed. To eliminate extreme observations, any results outside the $\pm 2\sigma$ range from the mean were excluded. After the elimination of the extreme data, means and standard deviations were re-calculated for each scene and for each observer. The mean values were used for the purpose of observer calibration.

From the sharpness matching experiment of the small-large pairs of images, we saw an average change of perceived sharpness of 7.07 JNDs with $\sigma = 0.897$. The range of changes was from 3.92 to 9.20 JNDs. 75% of the test images were within the range of the mean $\pm 1\sigma$. From the small-medium pair experiment, the change in perceived sharpness was 3.70 JNDs with $\sigma = 0.540$. The range of changes was from 2.56 to 5.20 JNDs. 73% of the test images were within the range of the mean $\pm 1\sigma$. From the medium-large experiment, the change was 3.01 JNDs with $\sigma = 0.544$. The range of changes was from 0.86 to 4.73 JNDs. 80% of the test images were within the range of the mean $\pm 1\sigma$.

The results from these psychophysical experiments clearly show that the perceived sharpness was affected by changes in the displayed image size. However, the changes in perceived sharpness were not proportional to the changes in displayed image size. The degree of change was larger for the small-medium pair experiment than for the medium-large pair experiment.

The mean appearance change (in JNDs) obtained from the small-large pair experiment was found to be almost the same as the sum of the mean appearance change from the small-medium and the medium-large experiments. The discrepancy was equal to 0.35 JNDs for a total range of 0 to 2.08 JNDs. This finding, not only validates the three different experiments, but implies that the results can possibly be interpolated for other image sizes.

The appearance changes in JNDs are shown for all images in Figure 8; they are arranged based on the degree of change of each image. Detailed data are presented in Appendix A.

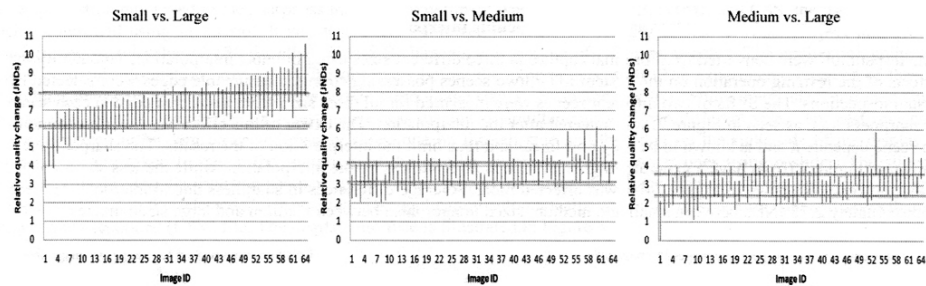


Figure 8. Evaluation of changes in relative image quality with respect to displayed image size.

4.3 Changes of MTF due to changes in angle of subtense

In parallel to the evaluation of changes in sharpness by variations in the displayed image size, the effect of changes in relative quality JNDs because of different the angle of subtense was computed. This is based on the assumption that the image obtained from the imaging system possessed equal quality to that of a diffraction limited lens. Changes in the visual angle were calculated by changing the spatial frequency (in visual degrees) of the displayed image size. The changes in JNDs with respect to changes in the visual angle are presented in Figure 9. They are shown to clearly non-linear. The quality was not much affected when the change in visual angle was less than 30%. Above 30%, there is a quality loss. Quality changes dramatically after 70% of change in visual angle.

The region that represents the small-large experiment was found at approximately 11 JNDs. Also, regions representing the small-medium and the medium-large experiments were found at approximately 5.66 and 5.34 JNDs, respectively. These results show a pattern similar to that in the results obtained from the sharpness matching experiments.

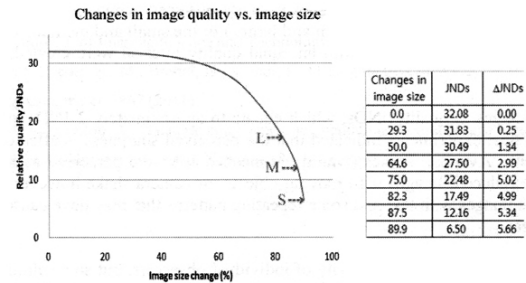


Figure 9. Changes in image quality with respect to the changes in angle of view.

4.4 Scene interpretation

The results in section 4.2 show that the degree of changes in perceived sharpness differed for each scene. In an effort to provide an interpretation for the variations in the results, the test scenes characteristics were investigated. Scene busyness has been identified to correlate with image sharpness in a number of studies [16, 17]. The busyness of the test images was evaluated using segmentation techniques suggested by Triantaphillidou *et al* [16] and by Gupta *et al* [18], but no strong correlations were found for the scenes outside of the average JND range.

In addition to this objective scene analysis, the test images were visually inspected. It was noticed that, the test images that were less affected with changes in displayed image size included either moving objects, or camera shake. So, they were rather unsharp to start with. The test images that were most affected by changes in image size included text, or repeated objects. Some examples are presented in Figure 10.



Figure 10. Least affected test images with changes in displayed image sizes (top), and most affected test images (bottom).

5. CONCLUSIONS

A series of psychophysical experiments were carried out to investigate the changes in perceived image sharpness when images are viewed at different displayed sizes on a high end LCD. For test image capture, a digital SLR equipped with a quality zoom lens was used. A total of 64 natural scenes with various scene contents were selected. The images were resized to generate three different sizes: small, medium and large. For the small and medium versions of the test images, a set of 25 images of varying image sharpness with an equal quality interval were created, using frequency domain filtering.

Results were provided as changes in quality JNDs, which equate to approximately 2 JNDs in perceived sharpness [9]. Results from the psychophysical experiments indicated that the perceived sharpness is affected by physical changes in image size and scene content. 'Average' scenes (within the mean $\pm 1\sigma$) were perceived as being less sharp when the displayed images sizes were reduced. Images with moving objects, or camera shake however appeared to be sharper at smaller image sizes, whereas images with text and some repeating patterns that may have caused aliasing appeared more blurred at smaller images sizes.

Observer variations in the results reflect the sensitivity of individual observers, but some similar tendencies were found within the data for each observer. Further scene analysis is ongoing in an attempt to identify the correlation between the degree of changes in perceived sharpness and the scene characteristics.

REFERENCES

- [1] Bartleson, C. J., Breneman, E. J., "Brightness Perception in Complex Fields", J. of Optical Society of America, 57(7), 953-956 (1967)
- [2] Choi, S. Y., Luo, M. R., Rhodes, P. A., Heo, E. G., Choi, I. S., "Colorimetric Characterization Model for Plasma Display Panel", J. of Imaging Science and Technology, 51(4), 337 (2007)
- [3] Nezamabadi, M., Berns, R. S., "The effect of image size on the color appearance of image reproductions using colorimetrically calibrated LCD and DLP displays", J. of Society for Information Display, 14(9), 773-783 (2005)
- [4] Nezamabadi, M., Montag, E. D., Berns, R. S., "An Investigation of the Effect of Image Size on the Color Appearance of Softcopy Reproductions using a Contrast Matching Technique", Proc. SPIE/IS&T Electronic Imaging 6493 (2007)
- [5] Xiao, K., Luo, M. R., Li, C. J., Rhodes, P. A., Taylor, C., "Specifying the Color Appearance of a Real Room", Proc. IS&T/SID Color Imaging Conference 2003, 308 (2003)
- [6] Park, J., Y., Triantaphillidou, S., Jacobson, R., E., "Identification of image attributes that are most affected with changes in displayed image size", Proc. SPIE/IS&T Electronic Imaging Conference 7242 (2009)
- [7] Wang, Z. H., Hardeberg, J. H., "Development of an Adaptive Bilateral Filter for Evaluating Color Image Difference", J. of Electronic Imaging, in print.
- [8] ISO 20462, "Photography-Psychophysical experimental methods for estimating image quality-Part-3: Quality ruler method", International Organization for Standardization, Geneva (2005)
- [9] Keelan, B. W., [Handbook of Image Quality: Characterization and Prediction], Marcel Dekker, INC., New York, USA, Chapter 6 (2002)
- [10] ISO 14524, "Photography-Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs)", International Organization for Standardization, Geneva (1999)
- [11] ISO 12233, "Photography-Electronic still picture cameras-Resolution measurements", International Organization for Standardization, Geneva (2000)
- [12] Imatest <http://www.imatest.com/>

- [13] Gonzalez, R. C., Woods, R. E., [Digital Image Processing, 2nd ed.], Prentice Hall, New Jersey, USA (2002)
- [14] Oakdale Engineering <http://www.oakdaleengr.com/>
- [15] Mozilla Firefox <http://www.mozilla.org/>
- [16] Triantaphillidou, S., Allen, E., Jacobson, R. E., "Image Quality Comparison Between JPEG and JPEG2000 II. Scene Dependency, Scene Analysis, and Classification", J. of Imaging Science and Technology, 51(3), 259-270 (2007)
- [17] Falkenstein, K., Bonnier, N., Pedersen, M., Brettel, H., Vienot, F., "Using metrics to assess the ICC perceptual rendering intent", Proc. SPIE Electronic Imaging Conference 7867 (2011)
- [18] Gupta, G., Psarrou, A., Angelopoulou, A., "Generic color image segmentation via multi-stage region merging", 10th International Workshop on Image Analysis for Multimedia Interactive Services, 185-188 (2009)

Just noticeable differences in perceived image contrast with changes in displayed image size

Jae Young Park*, Sophie Triantaphillidou and Ralph E. Jacobson
University of Westminster, Watford Road, Harrow, HA1 3TP, UK

ABSTRACT

An evaluation of the change in perceived image contrast with changes in displayed image size was carried out. This was achieved using data from four psychophysical investigations, which employed techniques to match the perceived contrast of displayed images of five different sizes. A total of twenty-four S-shape polynomial functions were created and applied to every original test image to produce images with different contrast levels. The objective contrast related to each function was evaluated from the gradient of the mid-section of the curve (gamma). The manipulation technique took into account published gamma differences that produced a just-noticeable-difference (JND) in perceived contrast. The filters were designed to achieve approximately half a JND, whilst keeping the mean image luminance unaltered. The processed images were then used as test series in a contrast matching experiment. Sixty-four natural scenes, with varying scene content acquired under various illumination conditions, were selected from a larger set captured for the purpose. Results showed that the degree of change in contrast between images of different sizes varied with scene content but was not as important as equivalent perceived changes in sharpness¹.

Keywords: Image quality, image appearance, perceived image contrast, image size, contrast matching, liquid crystal displays, LCDs, just-noticeable-differences, JNDs

1. INTRODUCTION

Changes in image size, or the viewing distance have been reported to lead to changes in various aspects of image appearance²⁻⁷. A study concerning the identification of image attributes that are most affected by changes in the displayed image size was previously carried out by the authors⁷. It considered various image attributes, including both spatial and color aspects and identified *sharpness* and *contrast* to be the two most affected attributes by changes in the displayed image size. Similar results were found in a recent study conducted by Wang *et al*⁸.

In a recent study¹, a series of psychophysical experiments were carried out to quantify changes in perceived sharpness with respect to changes in displayed image size. Results from the sharpness matching experiment showed that perceived sharpness increased when image size was decreased, but the magnitude of the perceived differences was scene dependent.

Here, first we investigated the effect of bi-cubic interpolation on image contrast by measuring root-mean-square (RMS) luminance contrast⁹⁻¹². No significant effect of bi-cubic interpolation on image contrast was evident. So we specifically focused our study on the quantification of changes in perceived global contrast with respect to changes in displayed image size. This was achieved by collecting data from psychophysical investigations that used techniques to match the perceived contrast of displayed images of five different sizes. The preparation of the test stimuli is presented in Section 2. Section 3 describes the psychophysical experiment and test conditions. Results are included and discussed in Section 4 and conclusions are drawn in Section 5.

* J.Park2@westminster.ac.uk; +44-(0)203-5068308

2. TEST STIMULI PREPARATION

2.1 Image capture

A large number of test images were acquired, using a Canon EOS 30D digital SLR camera, equipped with a CMOS sensor of 3504(h) x 2336(v) pixel resolution and an EF-S10-22mm zoom lens that provide a 35mm equivalent focal length of 16-35mm. A fixed focal length of 22mm was used for capturing images at different ISO settings and lens apertures. A total of sixty-four captured scenes were selected, after visual inspection of image quality and scene content. The selected scenes included architectural and natural scenes, portraits, artworks, and still and moving objects. They were recorded under various illumination conditions, had different original scene contrast and recorded noise levels, various amounts of fine detail, and strong lines and edges.

2.2 Creation of a series of filters for contrast manipulation with n-JND intervals

The contrast of reproduced scenes depends on the tone reproduction of the imaging systems employed. In display systems, tone reproduction is defined as the functional relationship between the input pixel values and the output luminance, and contrast can be expressed by gamma, γ . Bilissi *et al*¹³ have conducted various psychophysical experiments to evaluate acceptable and just perceptible gamma differences using cathode ray tube (CRT) displays under both controlled and uncontrolled environments, for a small image size, occupying 75 x 112mm of the faceplate area (corresponding to approximately 15 visual degrees when viewed from the viewing distance suggested in their paper). The just perceptible differences in gamma were 0.12 and 0.10 under controlled and uncontrolled environments, respectively.

The purpose of creating the filters was to produce test images with different contrast levels and thus enable us to quantify the changes in perceived image contrast with respect to changes in displayed image size. In this task, it was essential to take into account the perceptual gamma differences, whilst keeping the mean image luminance unaltered. Image manipulation using sigmoid functions to adjust image contrast has been used successfully in investigations in other laboratories^{14, 15} and was adopted for this work. This technique is based on the phenomenon of simultaneous lightness contrast. Thus, it is possible to make the highlight area in an image appear lighter by making the shadow area darker, which results in an increase in perceived image contrast.

A set of S-shaped filters, employed to increase image contrast and their corresponding inverse functions to decrease image contrast were created using the following steps. The step intervals were calculated by adjusting the gamma of the input to output transfer curve.

1. Pixel values (PV) ranging between 0 and 128 (half way the pixel values range) were selected and normalised (divided by 128).
2. Corresponding output PVs were calculated using a power function with exponent (gamma, γ), ranging between $\gamma = 1.6$ and $\gamma = 1/1.6$ with intervals of 0.05 gammas (approximately half a perceptible gamma difference).
3. Normalised original and corresponding PVs were reverted to their original range (0 to 128).
4. Corresponding output PVs were then mirrored at PV of 128 for the calculation of PVs between 128 and 255.
5. 6th order polynomials were fitted successfully to the calculated output pixel values using a curve fitting tool¹⁶.
6. Actual gammas of each filter function were derived for the mid-tones (linear) section of the functions.

Filter functions for the gamma adjustment are illustrated in Figure 1.

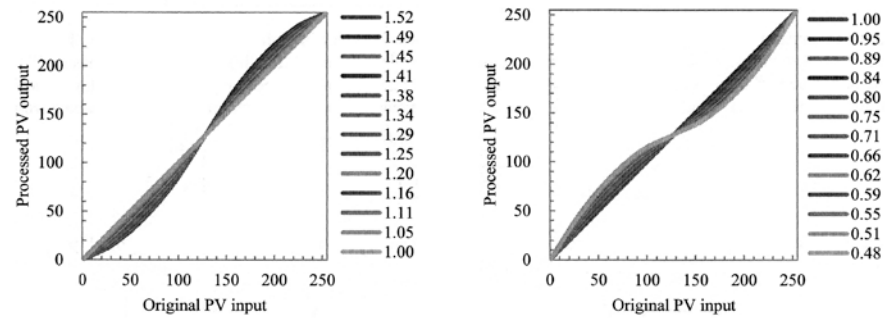


Figure 1. A series of gamma increasing filters (left) and gamma decreasing filters (right).

2.3 Contrast manipulation and bi-cubic interpolation

The filter operation was carried out using MATLAB. The filter functions were applied directly to the sixty-four original version images on the R, G, and B channels. A total of 25 ruler images, each possessing different contrast level with equal gamma difference (original, 12 contrast decreased versions, and 12 contrast increased versions), were generated in spatial domain. Sample image and its filtered versions were present with image histograms in Figure 2. The filtered images were then resized, using bi-cubic interpolation, to obtain five versions of the same scenes of different sizes. The test image dimensions were 744(h) x 560(v) pixels, 635(h) x 478(v), 526(h) x 396(v), 449(h) x 338(v), and 372(h) x 280(v) and represented large, large-medium, medium, medium-small and small sizes commonly displayed on computer and mobile device monitors. The small size was based on prevalent dimensions of the LCD on DSLR capturing devices. The large size was approximately half of the EIZO ColorEdge CG245W24.1"LCD's native horizontal and vertical resolution, which was later used for image appearance matching.

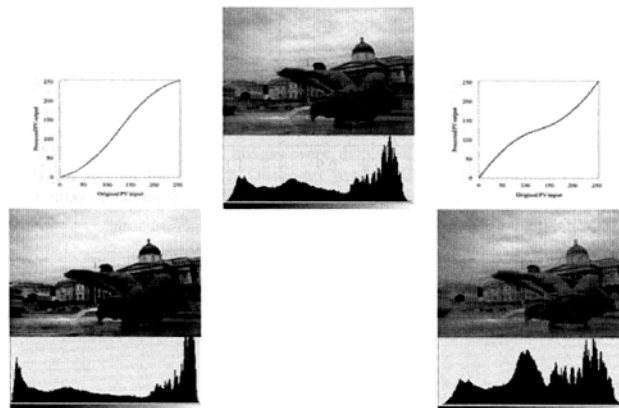


Figure 2. Sample S-shaped filters and the contrast manipulated images. Original image (top), contrast increased version at $\gamma = 1.52$ (bottom left) and contrast decreased version at $\gamma = 0.48$ (bottom right).

2.4 Objective contrast measurement of the ruler images

In order to confirm the changes in contrast of ruler images (i.e. test images with defined JNDs) objectively, the root mean square (RMS) contrast, which is one of the most commonly employed metrics for this purpose, was measured¹⁷. RMS contrast has been shown to correlate successfully with human contrast detection not only for the laboratory stimuli but also for natural images^{11, 18}. RMS contrast, C_{RMS} , of a two dimensional image is defined in Equation 1, adapted from Peli¹⁷.

$$C_{RMS} = \sqrt{\left[\frac{1}{R \cdot C} \sum_{x=1}^{C-1} \sum_{y=1}^{R-1} (I_{xy} - \bar{I})^2 \right]} \quad (1)$$

where R and C are the number of rows and columns in the image, I_{xy} is the normalised luminance of $x^{th}y^{th}$ pixel, \bar{I} is the mean normalised luminance of the image.

C_{RMS} of all sixty-four test images and that of their ruler versions were measured in display luminance space. Each original scene possessed a different C_{RMS} value and the degrees of change in C_{RMS} differed on ruler versions of each scene. However, changes in C_{RMS} on filtered images showed a linear trend. C_{RMS} values of four selected images of the large version are plotted in Figure 3 for illustration purposes. The selected scenes include those possessing the highest C_{RMS} and the lowest C_{RMS} and two scenes possessing average C_{RMS} .

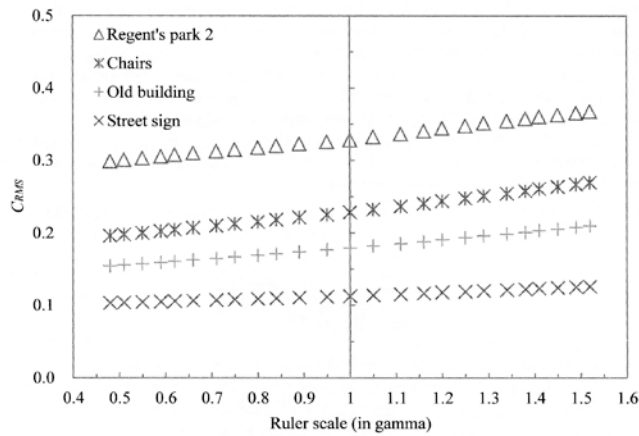


Figure 3. C_{RMS} of four selected scenes at a different ruler scale.

In addition, the effect of bi-cubic interpolation on the measured image contrast was investigated. C_{RMS} of all test images at five different sizes was measured. The effect of bi-cubic interpolation on C_{RMS} was not evident. C_{RMS} of the filtered 'Regent's Park 2' scene at various image sizes is shown in Figure 4.

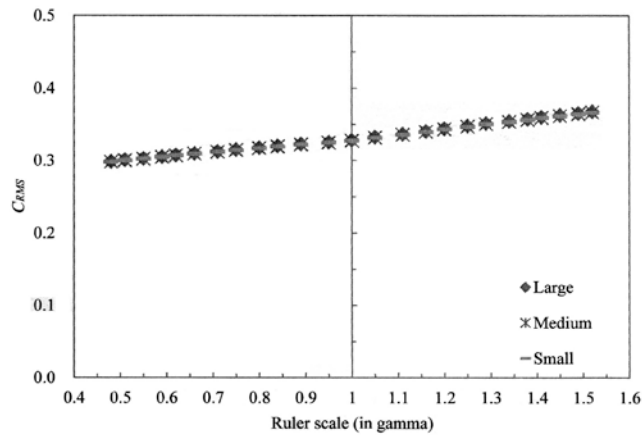


Figure 4. C_{RMS} of 'Regent's Park 2' at a different ruler scale in 3 different image sizes.

3. PSYCHOPHYSICAL INVESTIGATION

3.1 Calibration and settings of the system

The EIZO ColorEdge 245W LCD, driven by a Dell® Optiplex 760 computer with an ATI Radeon HD 3450 graphics controller, was used in the psychophysical investigation. The LCD has a native spatial resolution of 1,920 x 1,200 pixels and a tonal resolution of 24 bits (with a DVI connector). The system was set to a white point luminance of 120 cd/m², a gamma of 2.2 and a color temperature of D₆₅, using the GretagMacbeth Eye-One Pro with Profilemaker5. Daily calibration was carried out using the built-in calibration sensor.

3.2 Software preparation and interface design

The application employed in the contrast matching experiment was written in PHP, HTML and CSS, and the user interface was controlled using JavaScript. It was tested and optimized in Mozilla Firefox v5.0 web browser¹⁹. A mid-grey background in luminance (pixel value of R=G=B=186, at a gamma of 2.2) was selected. The application gathered some personal information provided by the observers before the experiments started. Each test image was displayed simultaneously in two different sizes. The test images were displayed in random order and display sides, adjacently placed on the left and right side of the display. Observers used a slider, controlled by the computer mouse, which simulated changes in the image contrast in response to changes in the slider position, by replacing the image frame with the appropriate ruler image. The display interface is illustrated in Figure 5.

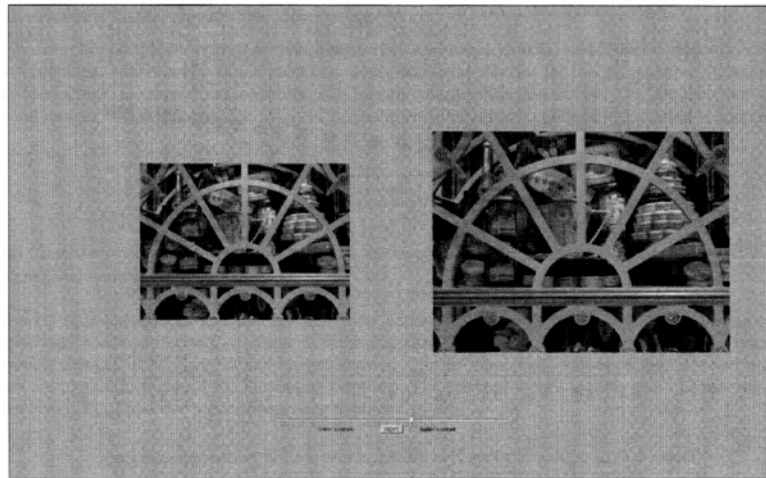


Figure 5. Graphical user interface of the contrast matching experiment.

3.3 Contrast matching

Contrast matching tests were conducted in a totally dark environment. Although the reference sRGB display viewing conditions were dim (ambient illuminance of 64 lux, and veiling glare of 0.2cd/m^2), the advantage of conducting experiments in such an environment was that the display was free from veiling glare, which is known to decrease perceived contrast and color saturation²⁰. Observers were seated on a comfortable seat with a chin rest to hold the observation distance at 60cm from the display and were requested to move their eyes from side to side only. During the tests, a randomly selected test image was displayed simultaneously at two different sizes on the LCD. The test images were displayed with random display position, one on the left side and the other on the right side of the display. The observers were asked to match the perceived image contrast of the smaller 'test' images to that of the larger 'standard' images using a slider. The application automatically wrote the observation data and saved them in a CSV file.

The experiment consisted of four matching contrast sessions: small image size to large image size, medium-small image size to large image size, medium image size to large image size, and large-medium image size to large image size. The observers completed the experiment in four separate sessions. Each observation took less than one hour per session; one session was conducted per day to avoid fatigue. A total of twenty observers, with normal visual acuity, participated. Their age ranged between 20 and 40 years old: all of them had imaging and design backgrounds.

4. RESULTS AND DISCUSSION

4.1 Results from the psychophysical tests

The mean, μ , and standard error of the mean (SEM) were calculated for each scene and size pairs. Results from the contrast matching experiment showed that perceived contrast was increased when image size was decreased and were consistent with results from the sharpness matching experiment⁷. Observations from matching the contrast of the small version image to that of the large version image, resulted in an average change in tone reproduction of 0.087 gamma (or 2.0 steps in the gamma scale), with an average standard error of mean (SEM) of 0.030. The range of change for all scenes was from -0.04 to 0.19. The changes for each scene, along with standard error, are plotted in Figure 6(a). From the experiment of the medium-small against large pairs of images, the average change was 0.050 gamma with an average SEM of 0.027. The range of change was from -0.08 to 0.14. From the medium version against the large version image, the average change was 0.043 gamma with an average SEM of 0.022. The range of change was from -0.02 to 0.13. And the large-medium version against the large version matching experiment showed that the average change was 0.036

gamma with an average SEM of 0.023. The range of change was from -0.054 to 0.096. The results are plotted in Figure 6 (b) to 6(d).

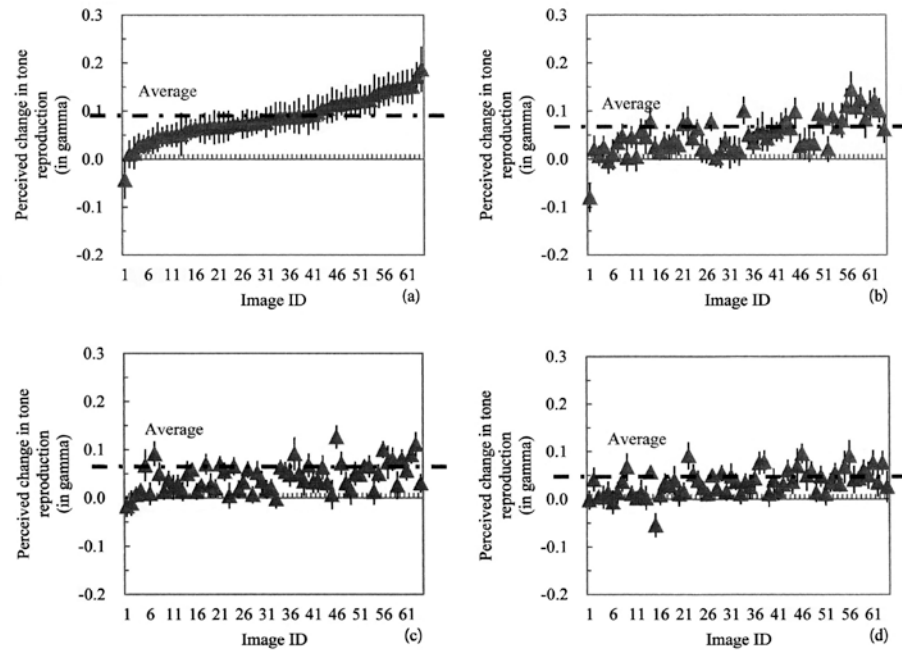


Figure 6. Average perceived change in tone reproduction (in gamma) with SEM. (a) small vs. large, (b) medium-small vs. large, (c) medium vs. large, (d) large-medium vs. large.

In addition to results in Figure 6, the average changes in perceived tone reproduction in gamma from all four experiments were plotted as a function of displayed image size in Figure 7. The figure clearly illustrates that the perceived contrast was proportionally affected by the changes in displayed image size. Smaller version images were perceived as having a higher contrast than that of the larger version. Similar results were found by recent research conducted by Haun *et al*²¹. The authors found that magnified video is perceived as having lower contrast than original video. Therefore, mirrored data at zero point has also been estimated by extrapolation and plotted as a linear function to predict change in perceived contrast when images may be displayed at larger scales. This assumes that the relationship remains linear. The linear trend line showed the relationship as: $y = -0.001x + 0.000$ with $R^2 = 0.983$.

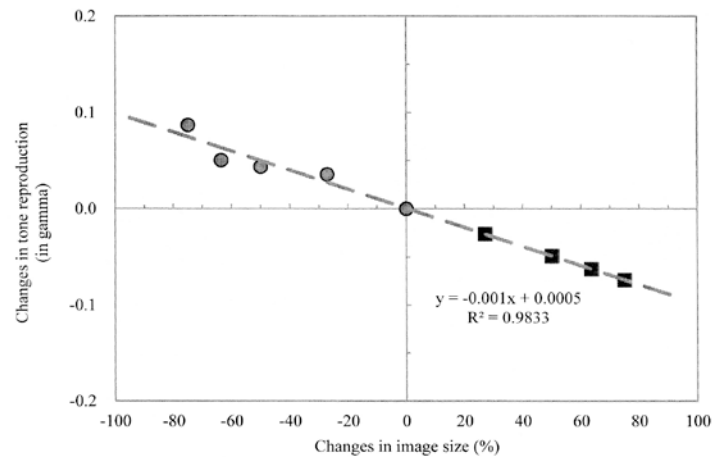


Figure 7. Perceived changes in tone reproduction with respect to the changes in displayed image size (blue circles) and extrapolated changes (gray squares) in non-calibrated relative image quality gamma scale.

4.2 Validation of the results

A psychophysical experiment to validate the results obtained from the contrast matching experiments was conducted under the environmental condition described in Section 3.3. A total of sixteen large size average scenes²² and their corresponding smaller versions, one unmodified and one contrast modified version according to previous finding, were used. Observers were asked to rate the test image pair in terms of their appearance matching (from 10 being the most matching to 1 being the least matching). A total of 7 average observers took part in the experiment.

Only in a few images, unmodified version images were perceived to be closer matching to the large original. However, majority of the contrast modified small versions were perceived to be closer matching to the large original than the unmodified small versions, as shown in Figure 8. The large original and contrast modified small version image pairs rated superior, $\mu = 4.90$, compared with the large original and small unmodified image pairs, $\mu = 4.62$.

However, in some cases the error bars overlap, making it unclear whether the contrast modified small images were clearly better than the unmodified originals or not. Nevertheless, the overall trends indicate that in general the contrast modification to compensate for image size modification is probably a worth-while operation, especially since it is not computationally expensive. Further work is needed to identify the original contrast characteristics of the scenes with close results.

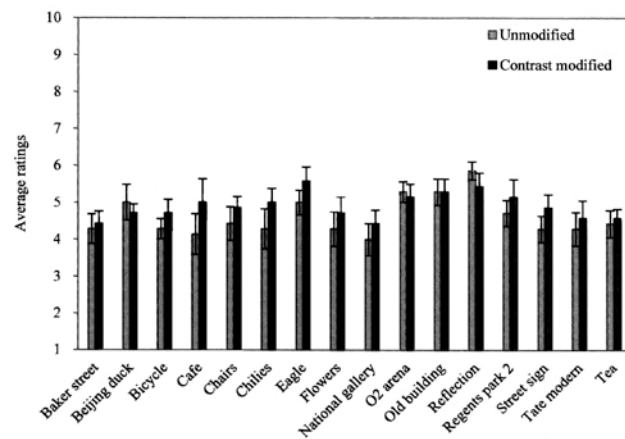


Figure 8. Average ratings for the contrast modified and unmodified image pairs.

4.3 Validation of step intervals and calibration in JND scale

A series of paired comparison experiments were conducted using all sixty-four scenes. For the step interval evaluation, only the central region of the scale (original ± 6 steps) was used, as most of the appearance changes were found within the range. A new set of filters with smaller intervals (half of the intervals used for contrast matching) was created for contrast enhancements to increase accuracy. Three male expert observers participated to a total of 192 sessions. Each observation took less than 10 minutes per session and a maximum of 10 sessions per day was conducted to avoid fatigue.

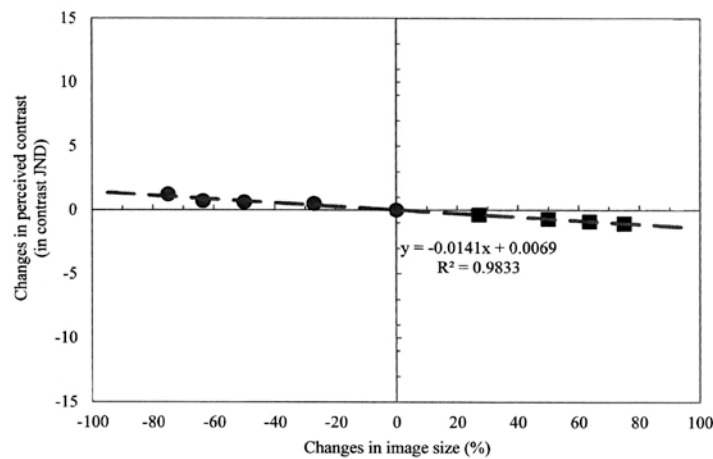


Figure 9. Changes in perceived contrast with respect to the changes in displayed image size (blue circles) and extrapolated changes (gray squares) in contrast JND scale.

An average of 0.070 gamma (or 1.614 steps) in the ruler scale was found to be 1 JND in perceived contrast. The outcome was used to calibrate the results obtained in Section 4.1 into contrast JND scales, plotted in Figure 9. The linear trend line shows the relationship as: $y = -0.014x + 0.006$ with $R^2 = 0.983$. The change in perceived contrast was approximately 1 JND with a 75% change in the displayed image size.

4.4 Changes in perceived contrast vs changes in perceived sharpness

Results from a sharpness matching experiment to define perceived sharpness changes with changes in displayed image size have been previously published¹. They are presented in Figure 10, calibrated in a sharpness JND scale²². The linear trend line shows the relationship as: $y = -0.159x - 0.069$ with $R^2 = 0.995$. Figures 9 and 10 allow a comparison between JNDs in sharpness changes versus JND in contrast changes, with similar changes in the display image size. The change in perceived sharpness was as much as 12 JNDs with a 75% change in the displayed image size, whereas the equivalent change in perceived contrast was 1 JND. Sharpness and contrast were previously identified to be the two most affected image attributes with respect to changes in displayed image size⁷, but perceived sharpness is shown in this study to be affected more severely compared to perceived contrast.

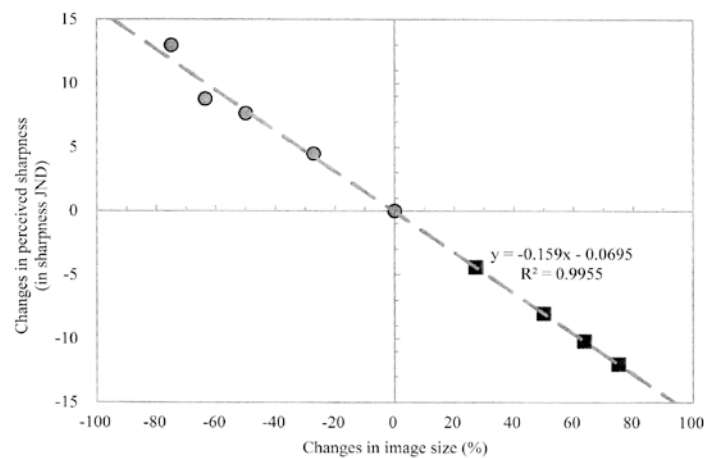


Figure 10. Changes in perceived sharpness with respect to the changes in displayed image size (blue circles) and extrapolated changes (gray squares) in contrast JND scale.

5. SUMMARY AND CONCLUSIONS

A series of psychophysical experiments were carried out to investigate the changes in perceived image contrast when images are viewed at different displayed sizes on an LCD device. A total of 64 natural scenes with various scene contents were selected for the purpose. The images were resized to generate five different sizes: small, medium-small, medium, large-medium, and large using bi-cubic interpolation. For the smaller versions of the test images, a set of 25 images of varying image contrast with an equal quality interval were created, using S-shape filters applied in the spatial domain.

For the range of image sizes we studied, no significant effect of the bi-cubic interpolation on image contrast was found. Results from the psychophysical matching experiments indicated that the perceived contrast was affected by changes in displayed image size. For the majority of the test images, smaller versions were judged as having a higher perceived contrast compared with the perceived contrast of the larger versions.

REFERENCES

- [1] Park, J. Y., Triantaphillidou, S., Jacobson, R. E. and Gupta, G., "Evaluation of perceived image sharpness with changes in the displayed image size," SPIE/IS&T Electronic Imaging, 82930J-82930J (2012).
- [2] Choi, S. Y., Luo, M. R. and Pointer, M. R., "Colour appearance change of a large size display under various illumination conditions," Electronic Imaging 2007, 649308-649308-9 (2007).
- [3] Nezamabadi, M. and Berns, R. S., "Effect of image size on the color appearance of image reproductions using colorimetrically calibrated LCD and DLP displays," Journal of the Society for Information Display 14(9), 773-783 (2006).
- [4] Nezamabadi, M., Montag, E. D. and Berns, R. S., "An investigation of the effect of image size on the color appearance of softcopy reproductions using a contrast matching technique," SPIE/IS&T Electronic Imaging, 6493091-6493099 (2007).
- [5] Xiao, K., Luo, M. R., Li, C. and Hong, G., "Colour appearance of room colours," Color Research & Application 35(4), 284-293 (2010).
- [6] Xiao, K., Luo, M. R., Li, C., Cui, G. and Park, D., "Investigation of colour size effect for colour appearance assessment," Color Research & Application 36(3), 201-209 (2011).
- [7] Park, J. Y., Triantaphillidou, S. and Jacobson, R. E., "Identification of image attributes that are most affected with changes in displayed image size," SPIE/IS&T Electronic Imaging, 724210 (2009).
- [8] Wang, Z. and Hardeberg, J. Y., "Development of an adaptive bilateral filter for evaluating color image difference," Journal of Electronic Imaging 21(2), 023021-1-023021-10 (2012).
- [9] Moulden, B., Kingdom, F. and Gatley, L. F., "The standard deviation of luminance as a metric for contrast in random dot images," Perception 19(1), 79-101 (1990).
- [10] Tiippana, K., Näsänen, R. and Rovamo, J., "Contrast matching of two-dimensional compound gratings," Vision Res. 34(9), 1157-1163 (1994).
- [11] Bex, P. J. and Makous, W., "Spatial frequency, phase, and the contrast of natural images," Journal of the Optical Society of America 19, 1096-1106 (2002).
- [12] Triantaphillidou, S., Jarvis, J. and Gupta, G., "Contrast sensitivity and discrimination of complex scenes," Proc. Image Quality and System Performance X, 8653, San Francisco, USA (2013), 86530C-86530C (2013).
- [13] Bilissi, E., Jacobson, R. E. and Attridge, G. G., "Just noticeable gamma differences and acceptability of sRGB images displayed on a CRT monitor," Imaging Science Journal, The 56(4), 189-200 (2008).
- [14] Braun, G. J. and Fairchild, M. D., "Image lightness rescaling using sigmoidal contrast enhancement functions," Journal of Electronic Imaging 8(4), 380-393 (1999).
- [15] Hassan, N. and Akamatsu, N., "A new approach for contrast enhancement using sigmoid function," the international arab journal of information technology 1(2), 221-225 (2004).
- [16] Oakdale Engineering, "Oakdale Engineering DataFit," 2007(11/25).
- [17] Peli, E., "Contrast in complex images," JOSA A 7(10), 2032-2040 (1990).
- [18] Frazor, R. A. and Geisler, W. S., "Local luminance and contrast in natural images," Vision Res. 46(10), 1585-1598 (2006).
- [19] Mozilla, "Mozilla Firefox," 2008(08/13) (2013).
- [20] Hunt, R. W. G., "Light and Dark Adaptation and the Perception of Color," J. Opt. Soc. Am. 42, 190-199 (1952).
- [21] Haun, A. M., Woods, R. L. and Peli, E., "Perceived contrast of electronically magnified video," Proc. SPIE, 21 (2011).
- [22] Park, J. Y., "Evaluation of changes in image appearance with changes in displayed image size," Ph.D. thesis, University of Westminster (2014).

Appendix D. List of abbreviations

Abbreviation	Description
CIE.....	International Commission on Illumination
CDP.....	Cycles per visual Degree
CIE76 ΔE_{ab}^*	Colour Difference Formula defined in 1976
CIEDE2000 ΔE_{00}^*	Colour Difference Formula defined in 2000
CIELAB.....	CIE 1976 $L^*a^*b^*$ colour space
CIELUV.....	CIE 1976 $L^*u^*v^*$ colour space
CIEXYZ.....	CIE 1931 XYZ tristimulus values
CPP.....	Cycles per Pixel
CRI.....	Colour Reproduction Index
CRT.....	Cathode Ray Tube
CSS.....	Cascading Style Sheets
CSV.....	Comma Separated Value
DCT.....	Discrete Cosine Transform
DFT.....	Discrete Fourier Transform
DP.....	Display Port
DVI.....	Digital Visual Interface
DWT.....	Discrete Wavelet Transform
EPIC.....	Effective Pictorial Information Capacity
EOTF.....	Electro-Optical Transfer Function
FIR.....	Finite Impulse Response

FOV.....	Field of View
FPN.....	Fixed Pattern Noise
FUN.....	Fidelity, Usefulness, and Naturalness
GUN.....	Genuineness, Usefulness, and Naturalness
HTML.....	Hyper Text Markup Language
HVS.....	Human Visual System
iCAM.....	Image Appearance Model
IPS.....	In-Plane Switching
IQM.....	Image Quality Metric
ISO.....	International Organization for Standardization
JND.....	Just Noticeable Difference
JPEG.....	Joint Photographic Experts Group
LCD.....	Liquid Crystal Display
LSF.....	Line Spread Function
MSE.....	Mean Square Error
MTF.....	Modulation Transfer Function
OECF.....	Opto-Electronic Conversion Function
PDF.....	Probability Density Function
PHP.....	Hypertext Preprocessor
PIC.....	Perceived Information Capacity
PSE.....	Point of Subjective Equality
PSF.....	Point Spread Function
PV.....	Pixel Value
RGB.....	Red, Green, Blue

RMS.....	Root Mean Square
RMSE.....	Root Mean Square Error
ROI.....	Region of Interest
s-CIELAB.....	Spatial extension of CIELAB
SEM.....	Standard Error of Mean
SFR.....	Signal Frequency Response
SNR.....	Signal to Noise Ratio
SQF.....	Subjective Quality Factor
SQRI.....	Square Root Integral
SQS.....	Standard Quality Scale
SQS ₂	Secondary Standard Quality Scale
sRGB.....	Standard RGB colour space
SSIM.....	Structural Similarity Index
SVG.....	Scalable Vector Graphics
TIFF.....	Tagged Image File Format
VC _{ab} [*]	Variance in chroma
VIF.....	Visual Information Fidelity

References

CIE Standard colorimetric observers. (1991). Geneva: BS ISO CIE. ISO 10527:1991.

Graphic technology and photography--Colour characterisation of digital still cameras (DSCs)--Part 1: Stimuli, metrology and test procedures. (2006). Geneva: International Organization for Standardization. ISO 17321-1:2006.

Graphic Technology: Prepress digital data exchange--CMYK standard color image data (CMYK/SCID). (1997). Geneva: International Organization for Standardization. ISO 12460:1997.

ILV: International Lighting Vocabulary. (1987). Vienna: Commission Internationale de L'eclairage. CIE S 17.4.

Imatest. (2013). [online] Imatest LCC. Available from: <<http://www.imatest.com/>> [Accessed 09/01 2013].

Information technology--Digital compression and coding of continuous-tone still images: Requirements and guidelines. (1994). ISO/IEC 10918-1: 1994 ed. Geneva: International Organization for Standardization. ISO/IEC 19018-1:1994.

Information technology--Office equipment--Measurement of image quality attributes for hardcopy output--Binary monochrome text and graphic images. (2001). Geneva: International Organization for Standardization. ISO/IEC 13660:2001.

Information technology--Office equipment--Test charts and methods for measuring monochrome printer resolution. (2001). Geneva: International Organization for Standardization. ISO/IEC 29112:2012.

Multimedia Systems and Equipment--Colour measurement and management--Part 2.1: Default RGB colour space--sRGB. (2000). London: British Standards. BS EN 61966-2-1:2000.

Multimedia Systems and Equipment--Colour measurement and management--Part 4: Equipment using liquid crystal display panels. (2000). London: British Standards. BS EN 61966-4:2000.

Photography--Electronic scanners for photographic images--Part 2: Film scanners. (2004). Geneva: International Organization for Standardization. ISO 16067-2:2004.

Photography--Electronic still picture cameras-Methods for measuring opto-electronic conversion functions (OECFs). (1999). Geneva: International Organization for Standardization. ISO 14524:1999.

Photography--Electronic still picture cameras--Resolution measurements. (1999). Geneva: International Organization for Standardization. ISO 12233:1999.

Photography--Electronic still picture cameras--Resolution measurements. (2000). Geneva: International Organization for Standardization. ISO 12233:2000.

Photography--Psychophysical experimental methods for estimating image quality--Part 1: Overview of psychophysical elements. (2005). Geneva: International Organization for Standardization. ISO 20462-1:2005.

Photography--Psychophysical experimental methods for estimating image quality--Part 2: Triplet comparison method. (2005). Geneva: International Organization for Standardization. ISO 20462-2:2005.

Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method. (2005). Geneva: International Organization for Standardization. ISO 20462-3:2005.

Photography--Psychophysical experimental methods for estimating image quality--Part 3: Quality ruler method. (2012). Geneva: International Organization for Standardization. ISO 20462-3:2012.

Using SFRplus Part 1. (2013). [online] Imatest LCC. Available from: http://www.imatest.com/docs/sfrplus_instructions/ [Accessed 10/02 2012].

Abrardo, A. *et al.*, (1996). Art-Works Colour Calibration by Using the VASARI Scanner, *Fourth Color Imaging Conference: Color Science, Systems, and Applications*. November 1996. IS&T, 1996, pp. 94.

Allen, E., (2011). Spatial image processing. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

Axford, N.R., (1988). Evaluation of the photographic image. In: Jacobson, R.E., Ray, S.F., Attridge, G.G., (eds.) *The Manual of photography*. 8th ed. ed. London: Focal press.

Bala, R., (2002). Device characterization. In: Sharma, G., (ed.) *Digital color imaging handbook*. Boca Raton: CRC Press.

Bang, Y.S. *et al.*, (2008). Printer resolution measurement based on slanted edge method, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 6808, Image Quality and System Performance V*. 28 January 2008. IS&T, 2008, pp. 680807.

Barten, P.G.J., (1990). Evaluation of subjective image quality with the square-root integral method. *Journal of Optical Society of America*. 7 (10), 2024-2031.

Barten, P.G.J., (1991). Evaluation of the effect of noise on subjective image quality, *Electronic Imaging'91, San Jose, CA*. SPIE, 1991, pp. 2-15.

Bartleson, C.J., (1982). Combined influence of sharpness and graininess on the quality of colour prints. *Journal of Photographic Science*. **30** 33-38.

Bartleson, C.J., (1975). Optimum image tone reproduction. *Journal of the Society of Motion Picture Engineers*. **84** (8), 613-618.

Bartleson, C.J. and Breneman, E.J., (1967a). Brightness perception in complex fields. *Journal of Optical Society of America*. **57** (7), 953-956.

Bartleson, C.J. and Breneman, E.J., (1967b). Brightness reproduction in the photographic process. *Photographic Science and Engineering*. **11** (4), 254-262.

Berns, R.S. *et al.*, (1993). CRT colorimetry. Part II: metrology. *Color Research and Application*. **18** (5), 315-325.

Berns, R.S. and Shyu, M.J., (1995). Colorimetric characterization of a desktop drum scanner using a spectral model. *Journal of Electronic Imaging*. **4** (4), 360-372.

Bex, P.J. and Makous, W., (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America*. **19** 1096-1106.

Bilissi, E., (2004). *Aspect of image quality and the internet*, Thesis (PhD) - University of Westminster.

Bilissi, E. *et al.*, (2008). Just noticeable gamma differences and acceptability of sRGB images displayed on a CRT monitor. *Imaging Science Journal, the*. **56** (4), 189-200.

Boring, E.G., (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts.

Boynton, R.M., (1984). Psychophysics. In: Bartleson, C.J. and Grum, F.C., (eds.) *Optical radiation measurements, Vol.5*. New York: Academic Press.

Boynton, R.M., (1961). Some temporal factors in vision. In: Rosenblith, W., (ed.) *Sensory communication*. New York: John Wiley.

Braun, G.J. and Fairchild, M.D., (1999). Image lightness rescaling using sigmoidal contrast enhancement functions. *Journal of Electronic Imaging*. **8** (4), 380-393.

Burns, P.D., (2011). Estimation error in image quality measurements, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7867, Image Quality and System Performance VIII*. 24 January 2011. SPIE, 2011, pp. 78670H.

Burns, P.D., (2000). Slanted-edge MTF for digital camera and scanner analysis, *Proc. IS&T PICS 2000: Image Processing, Image Quality, Image Capture, Systems Conference*. March 2000. IS&T, 2000, pp. 135.

Burns, P.D. and Williams, D., (2002). Refined Slanted-Edge Measurement for Practical Camera and Scanner Testing, *Proc. IS&T PICS 2002: Image Processing, Image Quality, Image Capture Systems Conference*. April 2002. IS&T, 2002, pp. 191.

Cambridge in colour, (2013a). *Digital camera image noise*. [online] Available from: <<http://www.cambridgeincolour.com/tutorials/image-noise.htm>> [Accessed 07/24 2013].

Cambridge in colour, (2013b). *Sharpness*. [online] Available from: <<http://www.cambridgeincolour.com/tutorials/sharpness.htm>> [Accessed 07/02 2013].

Cao, F. *et al.*, (2009). Measuring texture sharpness of a digital camera, In: B.G. Rodricks and S.E. Süssstrunk, eds. *Proc. SPIE 7250, Digital Photography V*. January 19, 2009. SPIE, 2009, pp. 72500H.

Cheung, T.L.V. and Westland, S., (2003). Accurate estimation of the non-linearity of input-output response for color digital cameras, *Proc. IS&T PICS 2003: Image Processing, Image Quality, Image Capture Systems Conference*. 13 May 2003. IS&T, 2003, pp. 366-369.

Cheung, V. *et al.*, (2004). Accurate estimation of the nonlinearity of input/output response for color cameras. *Color Research and Application*. **29** (6), 406-412.

Choi, S. *et al.*, (2007a). Colorimetric characterization model for plasma display panel. *Journal of Imaging Science*. **51** (4), 337-347.

Choi, S.Y. *et al.*, (2007b). Colour appearance change of a large size display under various illumination conditions, In: R. Eschbach and G.G. Marcu, eds. *Proc. SPIE 6493, Color Imaging XII*. 29 January 2007. SPIE, 2007b, pp. 649308.

Chou, C.H. *et al.*, (2008). P-86: Channel-Dependent GOG Model for Colorimetric Characterization of LCDs, *SID Symposium Digest of Technical Papers*. May 2008. SID, 2008, pp. 1510.

Dainty, J.C., (1974). *Image science: principles, analysis and evaluation of photographic-type imaging processes*. London: Academic Press.

Day, E.A. *et al.*, (2004). Colorimetric characterization of a computer-controlled liquid crystal display. *Color Research and Application*. **29** (5), 365-373.

Easton, R.L.J., (2010). *Fourier methods in imaging*. Chichester; Hoboken: John Wiley & Sons.

Engel drum, P.G., (2000). *Psychometric scaling: a toolkit for imaging systems development*. Winchester: Imcotek Press.

- Estribeau, M. and Magnan, P., (2004). Fast MTF measurement of CMOS imagers using ISO 12333 slanted-edge methodology, In: J.P. Chatard and P.N.J. Dennis, eds. *Proc. SPIE 5251, Detectors and Associated Signal Processing*. 19 February 2004. SPIE, 2004, pp. 243.
- Fairchild, M.D., (2005). *Color appearance models*. 2nd ed. Chichester; Hoboken: John Wiley & Sons.
- Fairchild, M.D., (1998). *Color Appearance Models*. Harlow: Addison-Wesley.
- Fairchild, M.D., (1995). Considering the surround in device-independent color imaging. *Color Research and Application*. **20** (6), 352-363.
- Fairchild, M.D. and Wyble, D.R., (1998). *Colorimetric characterization of the apple studio display (flat panel LCD)*. New York: Rochester Institute of Technology (RIT).
- Fairchild, M.D. and Johnson, G.M., (2004). iCAM framework for image appearance, differences, and quality. *Journal of Electronic Imaging*. **13** (1), 126-138.
- Falkenstern, K. *et al.*, (2011). Using metrics to assess the ICC perceptual rendering intent, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7867, Image Quality and System Performance VIII*. 24 January 2011. SPIE, 2011, pp. 786706.
- Farrell, J.E., (1999). Image quality evaluation. In: MacDonald, L. and Luo, M.R., (eds.) *Colour Imaging-Vision and Technology*. Chichester: John Wiley & Sons.
- Fechner, G.T., (1966). *Elements of Psychophysics, Translated by Helmut E. Alder*. New York: Holt, Rinehart and Winston.
- Ford, A.M., (1997). *Relationships between image quality and still image compression*, Thesis (PhD) - University of Westminster.
- Frazor, R.A. and Geisler, W.S., (2006). Local luminance and contrast in natural images. *Vision Research*. **46** (10), 1585-1598.
- Freiser, H. and Biedermann, K., (1992). Quality criteria. In: Baker, L.R., (ed.) *Selected papers on optical transfer function: measurement*. SPIE Press.
- Giorgianni, E.J. and Madden, T.E., (2008). *Digital color management: encoding solutions*. 2nd ed. Chichester: John Wiley & Sons.
- Glasser, J., (1997). Principles of display measurement and calibration. In: MacDonald, L.W. and Lowe, A.C., (eds.) *Display systems: design and applications*. Chichester: Wiley/SID.
- Gonzalez, R.C. and Woods, R.L., (2002). *Digital image processing*. 2nd ed. Upper Saddle River; London: Prentice Hall.

Granger, E.M. and Cupery, K.N., (1972). An optical merit function (SQF), which correlates with subjective image judgments. *Photographic Science and Engineering*. **16** (3), 221-230.

Hassan, N. and Akamatsu, N., (2004). A new approach for contrast enhancement using sigmoid function. *The International Arab Journal of Information Technology*. **1** (2), 221-225.

Heynacher, E. and Kober, F., (1976). *Resolving power and contrast*. Oberkochen: Carl Zeiss. 51.

Higgins, G.C., (1977). Image quality criteria. *Journal of Applied Photographic Engineering*. **3** (2), 53-60.

Higgins, G.C. and Wolfe, R.N., (1955). The relation of definition to sharpness and resolving power in a photographic system. *Journal of Optical Society of America*. **45** (2), 121-125.

Hunt, R.W.G., (1952). Light and Dark Adaptation and the Perception of Color. *Journal of Optical Society of America*. **42** 190-199.

Hunt, R.W.G., (2004). *The Reproduction of Colour*. 6th ed. Chichester: John Wiley & Sons.

Hunt, R.W.G., (1998). Why is black-and-white so important in colour? In: L. MacDonald, ed. *CIM'98 Colour Imaging and Multimedia*. 23 September. IS&T, 1998, pp. 54.

Hurter, F. and Driffield, V.C., (1890). Photochemical investigations and a new method of determination of the sensitiveness of photographic plates. *Journal of the Society of the Chemical Industry*. **9** 455-469.

Jacobson, R.E., (1993). Approaches to total quality for the assessment of imaging systems. *Information Services & use*. **13** (3), 235-246.

Jacobson, R.E. and Triantaphillidou, S., (2002). Metric approaches to image quality. In: MacDonald, L.W. and Luo, M.R., (eds.) *Colour image science: exploiting digital media*. Chichester: John Wiley & Sons.

Jain, A.K., (1989). *Fundamentals of Digital Image Processing*. New Jersey: Prentice Hall.

Jenkin, R.B., (2011a). Digital image processing in the frequency domain. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

Jenkin, R.B., (2011b). Images and image formation. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

- Jenkin, R.B., (2011c). Noise, sharpness, resolution and information. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.
- Jenkin, R.B. *et al.*, (2007). Effective pictorial information capacity as an image quality metric, In: L.C. Cui and Y. Miyake, eds. *Proc. SPIE 6494, Image Quality and System Performance IV*. 29 January 2007. SPIE, 2007, pp. 64940O.
- Jin, E.W. *et al.*, (2009). Softcopy quality ruler method: implementation and validation, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7242, Image Quality and System Performance VI*. 19 January 2009. SPIE, 2009, pp. 724206.
- Johnson, G.M. and Fairchild, M.D., (2000). Sharpness rules, *Proc of IS&T/SID 8 th Color Imaging Conference*. 7-10 November 2000. IS&T, 2000, pp. 24-30.
- Johnson, G.M. and Fairchild, M.D., (2002). Visual psychophysics and color appearance. In: Sharma, G., (ed.) *Digital color imaging handbook*. Boca Raton: CRC Press.
- Johnson, T., (1996). Methods for characterizing colour scanners and digital cameras. *Displays*. **16** (4), 183-191.
- Jones, L.A., (1931). On the Theory of Tone Reproduction, with a Graphic Method for the Solution of Problems. *Journal of the Society of Motion Picture Engineers*. **16** (5), 568-599.
- Jones, L.A., (1921). Photographic reproduction of tone. *Journal of Optical Society of America*. **5** 232.
- Jones, L.A. and Condit, H., (1941). The brightness scale of exterior scenes and the computation of correct photographic exposure. *Journal of Optical Society of America*. **31** (11), 651-678.
- Keelan, B.W., (2000). Characterization and Prediction of Image Quality, *IS&T PICS 2000*. IS&T, 2000, pp. 197.
- Keelan, B.W., (2002). *Handbook of Image Quality Characterization and Prediction*. New York: Marcel Dekker Inc.
- Kim, Y.T., (1997). Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Transactions on Consumer Electronics*. **43** (1), 1-8.
- Konica-Minolta, (2013). *Konica-Minolta CS-S10W*. [online] Available from: <<http://www.konicaminolta.jp/instruments/products/light/cs-s10w/>> [Accessed 09/12 2013].
- Koren, N., (2006). The Imatest program: comparing cameras with different amounts of sharpening, In: N. Sampat, J.M. DiCarlo, R.A. Martin, eds. *Proc. SPIE 6069, Digital Photography II*. 10 February 2006. IS&T, 2006, pp. 60690L.

Kuang, J. *et al.*, (2007). iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation*. **18** (5), 406-414.

Li, C. and Bovik, A.C., (2009). Three-component weighted structural similarity index, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7242, Image Quality and System Performance VI*. 19 January 2009. IS&T, 2009, pp. 72420Q.

Lockhead, G.R., (1992). Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences*. **15** (03), 543-558.

Luo, M.R. *et al.*, (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application*. **26** (5), 340.

Mancusi, F. *et al.*, (2010). Multidimensional image selection and classification system based on visual feature extraction and scaling, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7529, Image Quality and System Performance VII*. 18 January 2010. SPIE, 2010, pp. 75290A.

Maxwell, J.C., (1871). On colour vision. *Nature*. **4** 13-16.

McCamy, C.S. *et al.*, (1976). A color-rendition chart. *Journal of Applied Photographic Engineering*. **2** (3), 95-99.

McElvain, J. *et al.*, (2010). Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems, In: F. Imai, N. Sampat, F. Xiao, eds. *Proc. SPIE. 7537, Digital Photography VI*. 18 January 2010. SPIE, 2010, pp. 75370D.

Michelson, A.A., (1962). *Studies in optics*. Chicago: The University of Chicago Press.

Miyake, Y. *et al.*, (1984). Evaluation of image quality of digital pictures obtained by drum scanner and photo-printer. *Photographic and Electronic Image Quality*. 62-79.

Moulden, B. *et al.*, (1990). The standard deviation of luminance as a metric for contrast in random dot images. *Perception*. **19** (1), 79-101.

Mozilla, (2013). *Mozilla Firefox*. [online] Available from: <<http://www.mozilla.org/en-US/>> [Accessed 08/13 2008].

Nakamura, J., (2006). Basics of image sensors. In: Nakamura, J., (ed.) *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton: CRC Press.

Nelson, C.N., (1966). The theory of tone reproduction. *The Theory of Photographic Process*. 464-498.

- Nezamabadi, M. and Berns, R.S., (2006). Effect of image size on the color appearance of image reproductions using colorimetrically calibrated LCD and DLP displays. *Journal of the Society for Information Display*. **14** (9), 773-783.
- Nezamabadi, M. *et al.*, (2007). An investigation of the effect of image size on the color appearance of softcopy reproductions using a contrast matching technique, In: R. Eschbach and G.G. Marcu, eds. *Proc. SPIE 6493, Color Imaging XII*. 29 January 2007. SPIE, 2007, pp. 649309.
- Oakdale Engineering, (2007). *Oakdale Engineering DataFit*. [online] Oakdale Engineering. Available from: <<http://www.oakdaleengr.com/>> [Accessed 11/25 2007].
- Oh, K.H. *et al.*, (2010). Scene classification with respect to image quality measurements, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 7529, Image Quality and System Performance VII*. 18 January 2010. SPIE, 2010, pp. 752908.
- Orfanidou, M. *et al.*, (2008). Predicting image quality using a modular image difference model, In: S.P. Farnand and F. Gaykema, eds. *Proc. SPIE 6808, Image Quality and System Performance V*. 27 January 2008. SPIE, 2008, pp. 68080F.
- Park, J.Y. *et al.*, (2012). Evaluation of perceived image sharpness with changes in the displayed image size, In: F. Gaykema and P.D. Burns, eds. *Proc. SPIE 8293, Image Quality and System Performance IX*. 22 January 2012. SPIE, 2012, pp. 82930J.
- Pavel, M. *et al.*, (1987). Limits of visual communication: the effect of signal-to-noise ratio tin the intelligibility of American Sign Language. *Journal of Optical Society of America*. **4** (12), 2355-2365.
- Peli, E., (1990). Contrast in complex images. *Journal of Optical Society of America*. **7** (10), 2032-2040.
- Pierson, R.E. *et al.*, (1996). Evaluating effective resolution of an optical tomographic imaging system using a narrow-band correlation metric, *Proc. International Conference on Image Processing, 1996*. 16-19 September 1996. IEEE, 1996, pp. 535-538.
- Pointer, M.R. and Hunt, R.W.G., (1994). A Color Reproduction Index, *Second IS&T/SID Color Imaging Conference*. 1994, pp. 180-182.
- Poynton, C.A., (1996). *A Technical Introduction to Digital Video*. New York; Chichester: Wiley.
- Rasband, W., (2013). *National Institutes of Health, USA ImageJ V.1.42q*. [online] National Institutes of Health. Available from: <<http://rsbweb.nih.gov/ij/>> 2013].
- Reichenbach, S.E. *et al.*, (1991). Characterizing digital image acquisition devices. *Optical Engineering*. **30** (2), 170-177.

Roufs, J.A.J., (1989). Brightness contrast and sharpness, interactive factors in perceptual image quality, *OE/LASE'89, 15-20 Jan., Los Angeles. CA*. International Society for Optics and Photonics, 1989, pp. 66-72.

Sheikh, H.R. and Bovik, A.C., (2006). Image information and visual quality. *IEEE Transactions on Image Processing*. **15** (2), 430-444.

Sheikh, H.R. and Bovik, A.C., (2005). A visual information fidelity approach to video quality assessment, *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. 2005, pp. 23-25.

Sheikh, H.R. *et al.*, (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*. **14** (12), 2117-2128.

Stevens, S.S., (1951). Mathematics, measurement and psychophysics. In: Stevens, S.S., (ed.) *Handbook of experimental psychology*. New York: Wiley, pp.149.

Stevens, S.S., (1946). On the theory of scales of measurement. *Science*. **103** (2684), 677-680.

Stroebel, L.D. and Zakia, R.D., (1993). *The Focal encyclopedia of photography*. 3rd ed. Boston; London: Focal Press.

Thurstone, L.L., (1927). A law of comparative judgment. *Psychological Review*. **34** (4), 273-286.

Tiippana, K. *et al.*, (1994). Contrast matching of two-dimensional compound gratings. *Vision Research*. **34** (9), 1157.

Töpfer, K. and Jacobson, R.E., (1993). The relationship between objective and subjective image quality criteria. *Journal of Information Recording Materials*. **21** (1), 5-27.

Triantaphillidou, S., (2001). *Aspects of image quality in the digitisation of photographic collections*, Thesis (PhD) - University of Westminster.

Triantaphillidou, S., (2011a). Introduction to image quality and system performance. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

Triantaphillidou, S., (2011b). Tone reproduction. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

Triantaphillidou, S. and Allen, E., (2011). Digital image file formats. In: Allen, E. and Triantaphillidou, S., (eds.) *The manual of photography*. 10th ed. Oxford: Elsevier/Focal Press.

Triantaphillidou, S. *et al.*, (2007). Image Quality Comparison Between JPEG and JPEG2000. II. Scene Dependency, Scene Analysis, and Classification. *Journal of Imaging Science*. **51** (3), 259-270.

Triantaphillidou, S. *et al.*, (1999). An Evaluation of MTF Determination Methods for 35mm Film Scanners, *Proc. IS&T PICS 1999: Image Processing, Image Quality, Image Capture, Systems Conference*. 25 April 1999. IS&T, 1999, pp. 231-235.

Triantaphillidou, S. *et al.*, (2013). Contrast sensitivity and discrimination of complex scenes, In: P.D. Burns and S. Triantaphillidou, eds. *Proc. SPIE 8653, Image Quality and System Performance X*. 3 February 2013. SPIE, 2013, pp. 86530C.

Wang, Z. and Bovik, A.C., (2002). A universal image quality index. *IEEE Signal Processing Letters*. **9** (3), 81-84.

Wang, Z. *et al.*, (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transaction of Image Processing*. **13** (4), 600-612.

Wang, Z. and Hardeberg, J.Y., (2012). Development of an adaptive bilateral filter for evaluating color image difference. *Journal of Electronic Imaging*. **21** (2), 023021.

Wang, Z. *et al.*, (2003). Multiscale structural similarity for image quality assessment, *Proc. 37th IEEE Asilomar Conference on Signals, Systems and Computers*. 9-12 November. IEEE, 2003, pp. 1398.

Westland, S. *et al.*, (2012). *Computational Colour Science Using MATLAB*. 2nd ed. Chichester: John Wiley & Sons.

Williams, D. and Burns, P.D., (2014). Evolution of slanted edge gradient SFR measurement, In: S. Triantaphillidou and M.C. Larabi, eds. *Proc. SPIE 9016, Image Quality and System Performance XI*. 3 February 2014. SPIE, 2014, pp. 901605.

Xiao, K. *et al.*, (2004). Colour Appearance for Dissimilar Sizes, *Proc. Colour in Graphics, Imaging, and Vision 2004*. 1 January 2004. IS&T, 2004, pp. 12.

Xiao, K. *et al.*, (2011). Investigation of colour size effect for colour appearance assessment. *Color Research and Application*. **36** (3), 201-209.

Xiao, K. *et al.*, (2010). Colour appearance of room colours. *Color Research and Application*. **35** (4), 284-293.

Xiao, K. *et al.*, (2003). Specifying the Colour Appearance of a Real Room, *11th Color and Imaging Conference*. 1 January 2003. IS&T, 2003, pp. 308.

Yamada, T., (2006). CCD Imaging Sensors. In: Nakamura, J., (ed.) *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton: CRC Press.

Yendrikhovskij, S., (2002). Image quality and colour categorisation. In: Luo, M.R., (ed.) *Colour image science: exploiting digital media*. New York: John Wiley & Sons.

Yendrikhovskij, S., (1999). Image quality: Between science and fiction, *Proc. IS&T PICS 1999*. IS&T, 1999, pp. 173-178.

Yoshida, H., (2006). Evaluation of image quality. In: Nakamura, J., (ed.) *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton: CRC Press.

Young, T., (1802). The Bakerian lecture: On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*. **92** 12-48.

Zhang, X. *et al.*, (1997). Color image quality metric S-CIELAB and its application on halftone texture visibility, *Proc. IEEE Compcon'97*. 23-26 February 1997. IEEE, 1997, pp. 44.

Zhang, X. and Wandell, B.A., (1996). A spatial extension of CIELAB for digital color image reproduction, *SID International Symposium Digest Of Technical Papers*. SID, 1996, pp. 731-734.